



TAMPEREEN TEKNILLINEN YLIOPISTO

THUMAS MIILUMÄKI
WEB-POHJAISTEN SOSIAALISTEN VERKOSTOJEN
ANALYYSIMENETELMÄT

Diplomityö

Tarkastajat: Prof. Seppo Pohjolainen (TTY) ja
tutkija Jukka Huhtamäki (TTY)
Tarkastaja ja aihe hyväksytty
Luonnontieteiden ja ympäristötekniikan tiedekunta-
neuvoston kokouksessa 7. huhtikuuta 2010

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO

Teknis-luonnontieteellinen koulutusohjelma

THUMAS MIILUMÄKI: Web-pohjaisten sosiaalisten verkostojen

analyysimenetelmät

Diplomityö, 60 sivua, 22 liitesivua

Kesäkuu 2010

Pääaine: Matematiikka

Tarkastajat: Prof. Seppo Pohjolainen (TTY) ja tutkija Jukka Huhtamäki (TTY)

Avainsanat: SNA, graafiteoria, matriisilaskenta, sosiaalinen web

Viime vuosituhanen lopulla ja edelleen 2000-luvulla ihmisverkostojen tutkimuksen menetelmien joukosta on korostunut nk. SNA (Social Network Analysis) ja sen tarjoamat sovellukset. SNA kaikessa laajuudessaan tarjoaa monipuolisia menetelmiä verkostanalyysille tarkasteltaessa verkostoa niin yksittäisten toimijoiden kuin koko verkoston tasolla.

SNA:n matemaattinen teoreettinen tausta nojaa pääosiltaan kahteen matematiikan osa-alueeseen, graafiteoriaan ja matriisilaskentaan. Monessa kohtaa nämä kaksi teoria- ja metodologia-alueita kulkevat rinnakkain määriteltäessä SNA:n yleisiä teorioita ja metodeita verkostojen analysointiin. Graafiteoria tarjoaa perusperiaatteiltaan lähtökohdat verkoston visualisoinnille ja mallintamiselle. Matriisilaskenta puolestaan tarjoaa metodeita verkoston laskennalliselle analysoinnille.

Web-pohjaiset järjestelmät mahdollistavat laaja-alaisen ja yksityiskohtaisen lokidatan keruun ja tallentamisen. Keskeistä on kuitenkin huomioda, että toteutettavaa analyysiä ohjaavat hypoteesit, käytettävissä oleva lokidata, lokidatan rekonstruointi verkostodataksi sekä tiedon tulkinta osana analyysia. Koska web-pohjaiset sosiaaliset verkostopalvelut käsittävät laajoja verkostoja, ja siten verkostodata käsittää verrattain paljon tietoa, on järkevää määritellä verkostanalyysille johdonmukainen looginen prosessi, jota noudattamalla verkoston analyysi on paremmin vertailtavissa ja suhteutettavissa muihin tutkimustuloksiin.

Tässä diplomityössä esitellään niitä SNA-menetelmiä, jotka erityisesti soveltuvat web-pohjaisten sosiaalisten verkostojen analysointiin. Edelleen esitellään web-pohjaisen verkostoitumispalvelun SNA:n looginen prosessi, jota on hyödynnetty tukimuscasessa, jossa esiteltyjä menetelmiä testattiin käytännössä. Tutkimuksen tuloksena havaittiin, että tutkitun verkkopalvelun käyttöaktiivisuus ei ollut suoraan yhteydessä palvelun sisäiseen verkostoitumiseen.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme in Science and Engineering

THUMAS MIILUMÄKI : Social Network Analysis Methods in Web-based Networking Systems

Master of Science Thesis, 60 pages, 22 Appendix pages

June 2010

Major: Mathematics

Examiners: Prof. Seppo Pohjolainen (TUT) and researcher Jukka Huhtamäki (TUT)

Keywords: SNA, graph theory, matrix algebra, social web

In the end of the last millennium and, moreover, during the 21st century SNA (Social Network Analysis) methods and applications have been emphasised among human network research and study. SNA provides a wide range of methods for network analysis when studying a network from a single actor's point of view or in general level of the whole network.

Theoretical background of SNA is focused on two mathematical domains that are graph theory and matrix algebra. These two theory and methodology domains are in many ways parallel when defining general SNA theories. Graph theory offers basic principles for modelling and visualising social networks. Furthermore, theories of matrix algebra are used in quantitative analysis of networks.

Web-based systems enable to collect and store wide and precise log data. It is still crucial to notice that executed analysis is affected by hypotheses, available log data and reconstruction of the log data to network data as well as interpretation of knowledge as a part of analysis. Because Web-based social network sites are dealing with wide networks, and thus network data is rather complex, it is reasonable to define a logical process model for SNA. Using this process model network analysis is more comparable with other research results.

This Master of Science Thesis describes those SNA methods that are suitable for analysing Web-based social networks. Moreover, a logical process model for SNA of Web-based social networking site is presented and used in a case study. The case study was used to test the described theories in practice. As a result of this case study it was noticed that activity of the system usage was not directly affecting on networking of the system users inside the studied social networking site.

ALKUSANAT

Tämä työ on tehty sosiaalisten verkostojen tutkimusryhmässä, joka toimii Tampereen teknillisen yliopiston matematiikan laitoksella ja hypermedialaboratoriossa. Työn tutkimus-case liittyy Verkkoyhteisö opintojen aloittamisen tueksi -hankkeeseen.

Haluan kiittää työni tarkastajia professori Seppo Pohjolaista ja tutkija Jukka Huhtamäkeä, jotka ovat olleet keskeisesti ohjaamassa koko tutkimusryhmän työtä, mutta erityisesti tässä työssä käsiteltävää kvantitatiivista tutkimusmenetelmäpuolta.

Lisäksi haluan kiittää ryhmän muita jäseniä erikoistutkija Kirsi Siliusta ja tutkija Anne-Maritta Tervakaria perehdytyksestä kvalitatiiviseen tutkimukseen ja sen mahdollisuuksiin sosiaalisten verkostojen tutkimuksessa.

Lopuksi haluan vielä kiittää vanhempiani ja sisaruksiani, jotka ovat tukeneet minua koko opiskelujeni ajan, sekä Jaanaa, joka kannusti minua saattamaan tämän työn loppuun ja lopulliseen muotoonsa.

Tampere, 8. kesäkuuta 2010

Thumas Miilumäki
Tapettikatu 5 A 11
FI-33270 Tampere
thumas.miilumaki@tut.fi

SISÄLLYS

1. Johdanto	1
2. Teoreettinen tausta	3
2.1 Graafiteoria SNA:ssa	4
2.1.1 Sosiogrammit	4
2.1.2 Suuntaamattoman verkoston sosiogrammi	5
2.1.3 Suunnatun verkoston sosiogrammi	6
2.2 Graafien ominaisuuksia	7
2.2.1 Triviaali ja tyhjä graafi	7
2.2.2 Aligraafit	7
2.2.3 Solmun aste	9
2.2.4 Graafin tiheys	11
2.2.5 Kulku, reitti ja polku	12
2.2.6 Graafin yhtenäisyys	13
2.2.7 Isomorfiset graafit	16
2.2.8 Graafin komplementti	17
2.2.9 Puu ja metsä	17
2.2.10 Kaksijakoinen graafi	18
2.3 Matriisit SNA:ssa	18
2.3.1 Sosiomatriisi	19
2.3.2 Insidenssimatriisi	20
2.3.3 Muut SNA-matriisit	20
2.4 SNA-matriisien ominaisuuksia	21
2.4.1 Kulku	21
2.4.2 Saavutettavuus	21
2.4.3 Etäisyys ja geodeesi	23
2.4.4 Solmun aste	25
2.4.5 Tiheys	26
2.5 Arvotettujen verkostojen matriisit	27
2.6 SNA-tunnusluvut	27
2.6.1 Suuntaamattomien verkostojen tunnusluvut: Keskeisyys	28
2.6.2 Suunnattujen verkostojen tunnusluvut: Arvostus	30
2.7 Sosiaaliset verkostotyypit	33
2.7.1 Yksimoodinen verkosto	34
2.7.2 Kaksimoodinen verkosto	34
3. Web-pohjaiset sosiaaliset verkostopalvelut ja niiden SNA	36
3.1 Sosiaalisen verkostopalvelun SNA	36
3.2 Sosiaalisen verkostopalvelun analyysiprosessi	37

3.2.1	Verkostoanalyysin looginen prosessi	38
3.2.2	SNA-ohjelmistot	38
4.	Case: TTY-Piiri - Sosiaalisen verkostopalvelun verkostoanalyysi	40
4.1	TTY-Piiri	40
4.2	Tutkimuskysymykset	42
4.3	Interventiot ja käytönseuranta	43
4.4	Verkostodata	44
4.4.1	XML-formaatti	44
4.4.2	Pajek-formaatti	45
4.4.3	Matlab-formaatti	46
4.5	Analyysi	47
4.6	Visualisointi	54
4.7	Tulokset	55
4.8	Johtopäätökset ja pohdintaa	56
5.	Yhteenveto	58
	Lähteet	59
	A.Warshallin algoritmi: Matlab-koodi	61
	B.TTY-Piirin lokidata XML-formaatissa	62
	C.TTY-Piirin SNA-data Pajek-aikasarjaformaatissa	65
	D.TTY-Piirin SNA:n tulosten koonti	69
D.1	Alkutila: 31.8.2009	69
D.2	Välitila: 30.9.2009	71
D.3	Lopputila: 31.10.2009	77

KUVAT

2.1	Sosiogrammi: Suuntaamaton verkosto	5
2.2	Sosiogrammi: Suunnattu verkosto	6
2.3	Triadit	8
2.4	Irrotuspiste	15
2.5	Silta	15
2.6	Puu ja metsä	17
2.7	Kaksijakoinen digraafi	18
2.8	Kaksimoodisen verkoston graafi	34
3.1	Web-pohjaisen sosiaalisen verkoston analyysin looginen prosessi	38
4.1	TTY-Piirin ilmoitustaulu	41
4.2	TTY-Piirin käyttöaktiivisuus syksyllä 2009	43
4.3	TTY-Piirin automaattinen lokidatatyökalu.	44
4.4	TTY-Piirin alkutilan toimijoiden astelukujen jakauma	48
4.5	TTY-Piirin alkutilan toimijoiden normeerattujen keskeisyysasteiden jakauma	48
4.6	TTY-Piirin alkutilan toimijoiden normeerattujen läheisyyksien jakauma	49
4.7	TTY-Piirin alkutilan toimijoiden normeerattujen välillisyyksien ja- kauma	50
4.8	TTY-Piirin välitilan toimijoiden astelukujen jakauma	50
4.9	TTY-Piirin välitilan toimijoiden normeerattujen keskeisyysasteiden jakauma	51
4.10	TTY-Piirin välitilan toimijoiden normeerattujen läheisyyksien jakauma	51
4.11	TTY-Piirin välitilan toimijoiden normeerattujen välillisyyksien jakauma	52
4.12	TTY-Piirin lopputilan toimijoiden astelukujen jakauma	52
4.13	TTY-Piirin lopputilan toimijoiden normeerattujen keskeisyysasteiden jakauma	53
4.14	TTY-Piirin lopputilan toimijoiden normeerattujen läheisyyksien ja- kauma	54
4.15	TTY-Piirin lopputilan toimijoiden normeerattujen välillisyyksien ja- kauma	54
4.16	TTY-Piirin aikasarjadatan visualisointi	55

TAULUKOT

2.1	Tai- ja poissulkeva tai-konnektiivit: Totuustaulut	8
2.2	Sosiaalisen verkoston taulukkoesitys	19
2.3	Boolean summa	23

KÄYTETYT LYHENTEET

AJAX	Asynchronous JavaScript and XML
CMF	Content Management Framework
CMS	Content Management System
CMC	Computer-mediated Communication System
DNA	Dynamic Network Analysis
MySQL	My Structured Query Language
RSS	Really Simple Syndication
SNA	Social Network Analysis
XML	Extensible Markup Language
XOR	Exclusive or

TERMIT JA SYMBOLIT

\emptyset	Tyhjä joukko
$\{1, 2, 3, \dots\}$	Joukon alkiot
\forall	Kaikki/kaikille
\exists	On olemassa
\wedge	Looginen ja-konnektiivi
\vee	Looginen tai-konnektiivi
\oplus	Looginen poissulkeva tai-konnektiivi
$\binom{n}{k}$	Binomikerroin
$a!$	Kertoma $1 \cdot 2 \cdot 3 \cdots (a - 1) \cdot a$
$a \in A$	Alkio a kuuluu joukkoon A
$a \notin A$	Alkio a ei kuulu joukkoon A
$A \subseteq B$	Joukko A kuuluu joukkoon B tai on yhtenevä joukon B kanssa
(a, b)	Järjestämätön pari
$\langle a, b \rangle$	Järjestetty pari
A	Jäsenyysmatriisi tai hypergraafin matriisi
a_{ij}	Jäsenyysmatriisin alkio tai hypergraafin matriisin alkio
$card(\dots)$	Kardinaliteetti
C_B	Välillisyyden keskittyneisyys
$C_B(n_i)$	Solmun n_i välillisuus
$C'_B(n_i)$	Solmun n_i normeerattu välillisuus
$\overline{C_B}$	Normeerattujen välillisyyksien keskiarvo
C_C	Läheisyyden keskittyneisyys
$C_C(n_i)$	Solmun n_i läheisyys
$C'_C(n_i)$	Solmun n_i normeerattu läheisyys
$\overline{C_C}$	Normeerattujen läheisyyksien keskiarvo
C_D	Keskittyneisyysaste
$C_D(n_i)$	Solmun n_i keskeisyysaste
$C'_D(n_i)$	Solmun n_i normeerattu keskeisyysaste
$\overline{C_D}$	Normeerattujen keskeisyysasteiden keskiarvo
Δ	(Di)graafin tiheys

$d(n_i)$	Solmun n_i aste
$d_I(n_i)$	Solmun n_i tuontiluku
$d_O(n_i)$	Solmun n_i vientiluku
$d(i, j)$	Solmujen n_i ja n_j välinen etäisyys
\bar{d}	Graafin solmujen keskiaste
\bar{d}_I	Digraafin solmujen keskituontiluku
\bar{d}_O	Digraafin solmujen keskivientiluku
\mathcal{G}	Suuntaamaton graafi (sosiogrammi)
\mathcal{G}_d	Suunnattu graafi eli digraafi (sosiogrammi)
\mathcal{G}_s	Aligraafi
$\overline{\mathcal{G}}$	Graafin komplementti
$\overline{\mathcal{G}_d}$	Digraafin komplementti
\mathcal{H}	Hypergraafi
I_i	Toimijan vaikutusjoukon kardinaliteetti
\mathbf{I}	Insidenssimatriisi ja identiteettimatriisi
I_{ij}	Insidenssimatriisin alkio
K_g	Täydellinen graafi
K_{g^1, g^2}	Täydellinen kaksijakoinen graafi
$\kappa(\mathcal{G})$	Graafin \mathcal{G} piste- tai solmuyhtenäisyys
l_k	Kaari/nuoli k , $k = 1, 2, 3, \dots$
\mathcal{L}	Graafin kaarien joukko / Digraafin nuolien joukko
\mathcal{L}_s	Aligraafin kaarien/nuolien joukko
$\lambda(\mathcal{G})$	Graafin \mathcal{G} viiva- tai kaariyhtenäisyys
$\max(\dots)$	Maksimi (tai Boolean summa)
m_i	Tapahtumaa i kuvaava solmu
\mathcal{M}	Tapahtumia kuvaavien solmujen joukko / Hypergraafin kaarijoukko
M_j	Hypergraafin kaari j , $j = 1, 2, 3, \dots$
\mathbb{N}	Luonnollisten lukujen joukko $\{1, 2, 3, \dots\}$
n_i	Solmu i , $i = 1, 2, 3, \dots$
\mathcal{N}	Graafin/Digraafin solmujen joukko
\mathcal{N}_s	Aligraafin solmujen joukko
$P_D(n_i)$	Solmun n_i arvostusaste
$P'_D(n_i)$	Solmun n_i normeerattu arvostusaste
$P_P(n_i)$	Solmun n_i arvostusläheisyys
$\overline{P_P}$	Keskiarvostusläheisyys

$P_R(n_i)$	Solmun n_i arvoasema
S^2	Varianssi
$\Sigma_{i=1}^n i$	Summa $1 + 2 + 3 + \dots + n$
v_k	Yhteyden k arvo tai paino
W	Kulku
W^{-1}	Vastakulku
\mathbf{X}	Sosiomatriisi
x_{ij}	Sosiomatriisin alkio
$\mathbf{X}^{[R]}$	Saavutettavuusmatriisi
$x_{ij}^{[R]}$	Saavutettavuusmatriisin alkio
$\mathbf{X}^{[\Sigma]}$	Sosiomatriisin potenssimatriisien summamatriisi

1. JOHDANTO

Ihmisten välisiä suhteita ja niistä muodostuvia verkostoja on tutkittu eri näkökulmista aina käyttäytymistieteistä yhteiskuntatieteisiin saakka. Sosiaalisten suhteiden ja niiden merkityksen tutkimuksessa on viime aikoihin asti käytetty pääosin laadullisia tutkimusmenetelmiä. Määrällisen tutkimuksen osalta on rajoitettu erilaisten tilastollisten menetelmien käyttöön. Kiinnostus analysoida verkostojen rakennetta ja verkoston toimintaa on lisännyt tutkimusta verkostanalyysin menetelmien kehityksessä.

Viime vuosikymmenen lopulla ja edelleen 2000-luvulla ihmisverkostojen tutkimuksen menetelmien joukosta on korostunut nk. SNA (*Social Network Analysis*) ja sen tarjoamat sovellukset. SNA kaikessa laajuudessaan tarjoaa monipuolisia menetelmiä verkostanalyysille tarkasteltaessa verkostoa niin yksittäisten toimijoiden kuin koko verkoston tasolla. Laskennalliset menetelmät mahdollistavat toimijoita ja verkostoa kuvaavien tunnuslukujen määrittämisen, mikä yhdessä relevanttien visualisointien kanssa tuottaa verkostosta ja sen toimijoiden välisistä suhteista helposti ymmärrettävää tietoa.

SNA:n matemaattiset menetelmät pohjautuvat pääosin graafiteoriaan ja sen mahdollistaviin verkostanalyysin metodeihin. Graafiteoria antaa laajat mahdollisuudet ihmisverkostojen mallintamiseen ja visualisointiin. Verkostojen laskennallisessa analyysissä käytetään matriisilaskentaa ja matriiseihin perustuvia malleja. Graafiteoria yhdistettynä matriisilaskennan menetelmiin luovat perustan SNA:n monipuolisille verkostanalyysin metodeille.

Web-pohjaiset sosiaaliset verkostot luovat uudenlaisen ympäristön SNA-menetelmien soveltamiselle. Tekniset tietokantaratkaisut mahdollistavat laajan ja yksityiskohtaisen datankeruun. Käytännössä mistä tahansa palvelussa tapahtuvasta toiminnasta pystytään keräämään tarkkaa lokidataa, joka yksinkertaisimmillaankin sisältää verkostanalyysissä vaadittavan tiedon. Rikas lokidata luo mahdollisuudet yksityiskohtaiselle verkostanalyysille, jonka tuloksena saadaan syvempää tietoa verkoston toimijoista, verkoston rakenteesta ja sen sisäisestä toiminnasta.

Tässä diplomityössä perehdytään SNA:n menetelmiin, joiden avulla voidaan analysoida ja visualisoida yksilöiden muodostamia (*yksimoodisia*) verkostoja, joissa yhteydet muodostuvat aina ihmisparien välille. Edelleen perehdytään menetelmiin, jotka mahdollistavat verkostojen, joiden edustajat voidaan jakaa kahteen eri tyyppiin

(*kaksimoodinen verkosto*) (esim. ihmiset ja tapahtumat), analyysin. Luvussa 2 esitetyt teorit perustuvat teoksessa Wasserman ja Faust (1994) *Social network analysis: Methods and applications* esitettyihin teorioihin, mikäli muuta ei ole mainittu. Muista lähteistä mainittakoon Scott (2000) ja Knoke ja Yang (2008), jotka ovat olleet tukena teorioiden tarkentamisessa ja laajentamisessa. Esitettyjä teorioita ja menetelmiä tarkastellaan erityisesti suhteessa niiden sovellettavuuteen Web-pohjaisten verkostojen analyysissä.

Lisäksi esitellään Tampereen teknillisen yliopiston Hypermedialaboratorion Verkoyhteisö opintojen aloittamisen tueksi -hankkeessa toteutetun sosiaalisen verkostoitumispalvelun toimijaverkostoja käsittelevä SNA menetelmiin pohjautuva case-tutkimus ja sen tulokset. Tutkimuskysymyksenä esitettiin, oliko tutkittavan palvelun käyttöaktiivisuudella suoraa vaikutusta palvelun sisäiseen verkostoitumiseen. Tähän liittyi myös ajatus interventioista palvelun sisäisen toiminnan ohjauksen välineinä.

Web-pohjaiseen verkostoitumispalveluun toteutettiin monipuolinen lokidatan keruutyökalu, jonka tuottamasta tiedosta monipuolinen verkostanalyysi oli mahdollista toteuttaa. Analyysi toteutettiin noudattaen tarkoin määriteltyä loogista prosessia, jonka myötä saatujen kvantitatiivisten SNA-tulosten avulla kyettiin vastaamaan tutkimuskysymykseen. Jatkossa kokemusten perusteella kehitettävien automatisoitujen SNA-metodien ja -välineiden avulla monipuoliset visualisoinnit yhdessä rikkaan SNA-informaation kanssa ovat tuotettavissa niin tutkijan, kuin verkkopalvelun käyttäjänkin tarkasteltavaksi.

Tutkimus on osa ajankohtaista akateemista tutkimusta, josta on julkaistu kansainvälisiä tutkimusartikkeleja. Tämän diplomityön tekijä on ollut myös näiden akateemisten julkaisujen kirjoittajien joukossa. Julkaistut artikkelit ovat keskeisessä roolissa tässä esiteltävässä case-tutkimuksessa.

2. TEOREETTINEN TAUSTA

SNA:n matemaattinen teoreettinen tausta nojaa pääosiltaan kahteen matematiikan osa-alueeseen, graafiteoriaan ja matriisilaskentaan. Monessa kohtaa nämä kaksi teoria- ja metodologia-alueita kulkevat rinnakkain määriteltäessä SNA:n yleisiä määritelmiä ja metodeita verkostojen analysointiin. Graafiteoria tarjoaa perusperiaatteiltaan lähtökohdat verkoston visualisoinnille ja mallintamiselle. Matriisilaskenta puolestaan tarjoaa metodeita verkoston laskennalliselle analysoinnille. Molemmat teorialat mahdollistavat verkoston mallintamisen siten, että graafiteoreettinen malli on esitettävissä matriisina ja matriisimalli puolestaan graafina, mikä osaltaan yhdenmukaistaa SNA:n teoriapohjaa ja laajentaa sen tulkittavuutta. Carrington, Scott, ja Wasserman (2007) käsittelevät teoksessaan tarkemmin SNA:n malleja ja metodeita yleisesti ja yksittäistapauksissa sekä ottavat kantaa tulkinnan merkitykseen eri tilanteissa.

Tulkinnalla on erityinen merkitys verkostanalyysissä. Tulkinta vaikuttaa merkittävästi verkostodatan keruuseen ja sen sisältöön. Usein on järkevää asettaa ensin tarkka tutkimuskysymys, johon SNA-menetelmin pyritään vastaamaan. Tämän jälkeen määritellään kerättävä verkostodata, sen sisältö ja formaatti. Analyysin myötä tuotettu informaatio ei kuitenkaan aina itsessään selitä kaikkea sitä, mihin tutkimuksessa pyritään vastaamaan. Siksi onkin huomattava, että verkostanalyysin kvantitatiivinen puoli vaatii usein kvalitatiivista tutkimusta rinnalleen. Verkostanalyysissä kvantitatiiviset tutkimustulokset on monesti perusteltavissa sisällöllisen analyysin kautta. Hyvin toteutetussa verkostanalyysissä määrällinen ja laadullinen tutkimus siis täydentävät toisiaan.

Seuraavassa esitellään verkostanalyysin keskeisimpiä kvantitatiivisia menetelmiä ja niiden taustateorioita. Laadullisen analyysin ja tutkimuksen perusteita ei sen sijaan käsitellä. Verkostanalyysissä käytettävät kvalitatiiviset tutkimusmenetelmät riippuvat tutkittavasta kohteesta ja ilmiöstä. Edelleen kvalitatiivista tutkimusta on vaikea määritellä selvästi, koska sillä ei ole teoriaa eikä paradigmaa, joka olisi vain sen omaa. Kvalitatiivisella tutkimuksella ei myöskään ole täysin omia metodeja (Metsämuuronen, 2003, 162). Edelleen kvalitatiivinen tutkimus tarkoittaa eri asioita eri tutkimusaloilla (Denzin & Lincoln, 2005, 3). Mm. verkostanalyysissä käyttökelpoisia laadullisia tutkimusmetodeita ovat esittäneet teoksissaan esim. Eskola ja Suoranta (2000) ja Metsämuuronen (2003).

2.1 Graafiteoria SNA:ssa

Tapoja sosiaalisten verkostojen mallintamiseen on useita. Käytettävät menetelmät riippuvat mallinnuksen kohteesta, näkökulmasta ja tavoitteesta. Esimerkiksi monet tilastolliset mallinnusmenetelmät ovat sovellettavissa sosiaalisten verkostojen mallinnuksessa, mikäli verkostodata on valittuun tarkoitukseen relevanttia. Edelleen monet diagrammit ja kaaviot ovat sovellettavissa sosiaalisen verkoston datan käsittelyyn. Graafeilla pyritään luomaan mahdollisimman todenmukainen visualisointi, ikään kuin valokuva, tarkasteltavasta verkostosta, jossa toimijat ja niiden väliset yhteydet on esitetty selkeästi. Graafin ja todellisuuden yhteensovittaminen on haasteellista, sillä graafi luodaan usein diskreetistä ja rajallisesta määrästä verkostodataa. Freeman (2009) käsittelee artikkelissaan laajasti SNA:n visualisoinnin menetelmiä ja sen haasteita SNA-teorioiden kautta.

SNA:ssa graafien käytölle on useita perusteita. Ensinnäkin graafiteoreettinen sanasto soveltuu sosiaalisten rakenteiden kuvaamiseen ja merkitsemiseen erityisen hyvin. Toiseksi useimmat sosiaalisten rakenteiden ja ominaisuuksien kvantitatiiviset tunnusluvut ovat laskettavissa graafiteorian sisältämien matemaattisten menetelmien avulla. Kolmanneksi graafiteoreettinen notaatio ja sanasto sekä sen sisältämä matematiikka tarjoavat mahdollisuuden erilaisten graafeja koskevien teoreemien todistamiseen, ja siten myös sosiaalisia rakenteita koskevat väitteet ovat todistettavissa.

Täten graafiteoria tarjoaa mahdollisuuden sosiaalisen verkoston mallintamiseen. Graafi on esitys tarkasteltavasta verkosta, jossa reaali maailman toimijat esitetään graafissa solmuina / (kärki)pisteinä (*a node, nodes / a vertex, vertices*) ja niiden väliset yhteydet solmuja yhdistävinä kaarina (*an edge, edges*) tai nuolina (*an arc, arcs*).

2.1.1 Sosiogrammit

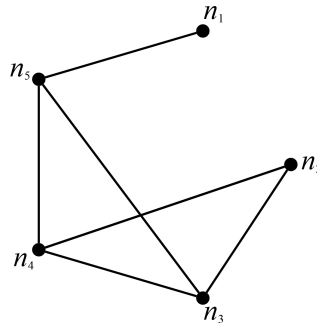
Jacob Levy Moreno esitteli jo 1930-luvulla sosiogrammin (*sociogram*), (Moreno, 1953, 96) joka on edelleen pohjana graafiteoreettisessa SNA:ssa ja verkostojen visualisoinnissa. Graafiteoria on ollut voimakkaasti käytössä eri tieteenaloilla mm. antropologiassa, kommunikaatiotutkimuksessa, elinkeino- ja liiketaloustutkimuksessa, organisaatiotutkimuksessa sekä maantieteissä. Edelleen graafiteorian tarjoamat menetelmät on otettu käyttöön piiriteoriassa ja -analyysissa. Lähes kaikilla em. tieteen- ja tutkimuksenaloilla on aina jollain tapaa ja jollain asteella mukana ihmisten muodostama sosiaalinen verkosto, jota graafiteoreettisin menetelmin on hyvä lähestyä ja analysoida.

Yleensä SNA:ssa käsitellään kahden tyyppisiä graafeja (tai sosiogrammeja) riippuen verkoston toimijoiden välisten suhteiden tyypistä. Verkostossa voidaan olettaa,

että toimijoiden välisellä yhteydellä joko on suunta tai suunta puuttuu (tai sitä ei huomioida). Käytännössä siis käsitellään verkostoja, joissa yhteydet ovat suuntaamattomia (*undirected*) tai suunnattuja (*directed*).

2.1.2 Suuntaamattoman verkoston sosiogrammi

Suuntaamattomassa graafissa verkoston toimijoiden väliset yhteydet ovat aina kaksisuuntaisia. Suuntaamaton graafi \mathcal{G} koostuu kahdesta joukosta; verkoston toimijoita kuvaavasta solmujen joukosta $\mathcal{N} = \{n_1, n_2, \dots, n_g\}$ (*a set of nodes*) ja yhteyksiä kuvaavasta kaarien (viivojen) joukosta $\mathcal{L} = \{l_1, l_2, \dots, l_L\}$ (*a set of lines*). Täten graafissa $\mathcal{G}(\mathcal{N}, \mathcal{L})$ on yhteensä g solmua ja solmuja yhdistäviä kaaria yhteensä L kappaletta.



Kuva 2.1: Suuntaamattoman verkoston graafi.

Määritelmä 2.1.1. Suuntaamattomassa graafissa jokainen kaari l_k on kahden erillisen solmun n_i ja n_j ei-järjestetty pari, ts.

$$l_k = (n_i, n_j) = (n_j, n_i).$$

Kaarta tai nuolta (suunnatuissa verkostoissa), jonka alkua- ja päätepisteenä on yksi ja sama solmu n_i , sanotaan silmukaksi (*a loop*) tai sisä-/refleksiivisidokseksi (*a reflexive tie*). Silmukoita ei useinkaan käytetä sosiogrammeissa, vaan sosiogrammit ovat silmukattomia, yksinkertaisia (*simple*), graafeja.

Lause 2.1.2. Oletetaan, että suuntaamattomassa yksinkertaisessa graafissa (sosiogrammissa) $\mathcal{G}(\mathcal{N}, \mathcal{L})$ on g solmua ($\text{card}(\mathcal{N}) = g$). Tällöin

$$\max(\text{card}(\mathcal{L})) = \frac{g(g-1)}{2}.$$

Todistus. Suuntaamattomassa graafissa, jossa on yhteensä g solmua, graafin jokainen solmu voi olla yhteydessä yhteensä $g-1$ solmuun, jolloin graafin kaikkien mahdollisten kaarien maksimimäärä on $g(g-1)$. Tällöin kuitenkin kutakin solmuparia yhdistäisi kaksi kaarta (jokainen on yhteydessä kaikkiin muihin). Siksi suuntaamattomassa graafissa voikin esiintyä vain puolet $g(g-1)$ kaaresta, eli maksimissaan $(g(g-1))/2$

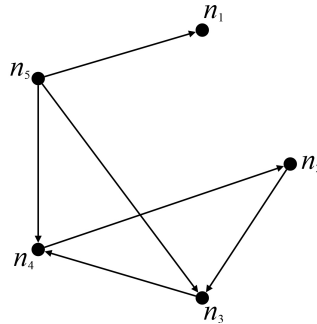
kaarta. □

Määritelmä 2.1.3. Kaksi solmua n_i ja n_j ovat vierekkäiset (*adjacent*), jos kaari $l_k = (n_i, n_j) \in \mathcal{L}$.

Määritelmä 2.1.4. Solmu n_i on liittynyt (*incident*) kaareen l_k , jos se on toinen solmuista, jotka muodostavat kaaren l_k määrittelevän ei-järjestetyn parin $l_k = (n_i, n_j)$.

2.1.3 Suunnatun verkoston sosiogrammi

Usein yhteydet voidaan tulkita suunnattuina, jolloin toimijoiden välisen yhteyden koetaan olevan orientoitunut toimijasta toiseen. Käsitteet vienti ja tuonti toimijoiden välisinä suhteina eroavat juurikin suuntansa puolesta toisistaan. Suunnatun verkoston graafi on digraafi (*digraph*). Digraafi \mathcal{G}_d koostuu kahdesta joukosta; verkoston toimijoita kuvaavasta solmujen joukosta $\mathcal{N} = \{n_1, n_2, \dots, n_g\}$ ja yhteyksiä kuvaavasta nuolien joukosta $\mathcal{L} = \{l_1, l_2, \dots, l_L\}$ (*a set of arcs*). Täten digraafissa $\mathcal{G}_d(\mathcal{N}, \mathcal{L})$ on yhteensä g solmua ja solmuja yhdistäviä nuolia yhteensä L kappaletta.



Kuva 2.2: Suunnatun verkoston graafi.

Määritelmä 2.1.5. Digraafissa jokainen nuoli l_k on kahden erillisen solmun n_i ja n_j järjestetty pari, ts.

$$l_k = \langle n_i, n_j \rangle \neq \langle n_j, n_i \rangle.$$

Nuoli $\langle n_i, n_j \rangle$ suuntautuu alku- tai lähettäjäsolmusta (*the origin or sender*) n_i päätte- tai vastaanottajasolmuun (*the terminus or receiver*) n_j .

Lause 2.1.6. Oletetaan, että yksinkertaisessa digraafissa (suunnatussa sosiogrammissa)

$\mathcal{G}_d(\mathcal{N}, \mathcal{L})$ on g solmua ($\text{card}(\mathcal{N}) = g$). Tällöin

$$\max(\text{card}(\mathcal{L})) = g(g - 1).$$

Todistus. Vastaavasti kuin suuntaamattoman graafin kohdalla, digraafista, jossa on yhteensä g solmua, jokainen digraafin solmu voi olla yhteydessä yhteensä $g-1$ solmuun, jolloin graafin kaikkien mahdollisten nuolien maksimimäärä on $g(g-1)$. Tällöin jokaista solmuparia yhdistää vastakkaissuuntainen nuolipari. \square

Määritelmä 2.1.7. Digraafin solmu n_i on vierekkäinen solmulle (*adjacent to*) n_j , jos $l_k = \langle n_i, n_j \rangle \in \mathcal{L}$.

Vastaavasti digraafin solmu n_j on vierekkäinen solmulta (*adjacent from*) n_i , jos $l_k = \langle n_i, n_j \rangle \in \mathcal{L}$.

Määritelmä 2.1.8. Solmu n_i on liittynyt nuoleen l_k , jos se on toinen solmuista, jotka muodostavat nuolen l_k määrittelevän järjestetyn parin $l_k = (n_i, n_j)$.

2.2 Graafien ominaisuuksia

2.2.1 Triviaali ja tyhjä graafi

Graafia $\mathcal{G}(\mathcal{N}, \mathcal{L})$, jossa on vain yksi solmu, sanotaan triviaaliksi (*trivial*). Tällöin siis

$$\mathcal{N} = \{n\}, \text{card}(\mathcal{N}) = 1, \mathcal{L} = \emptyset.$$

Tyhjässä (*empty*) graafissa ei ole yhtään kaarta solmujen välissä, ts.

$$\mathcal{N} = \{n_1, n_2, \dots, n_g\}, \mathcal{L} = \emptyset.$$

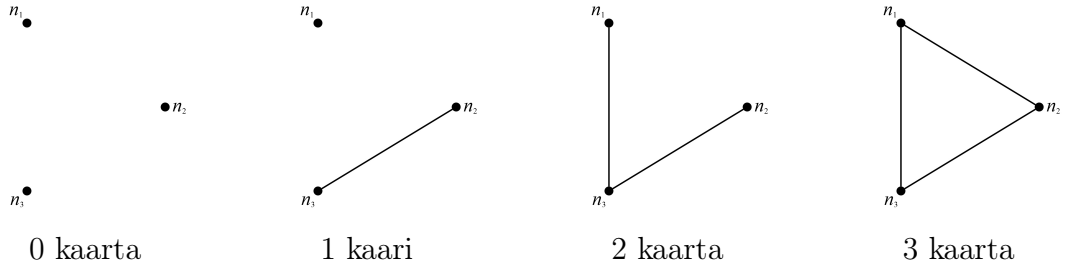
2.2.2 Aligraafit

Määritelmä 2.2.1. Graafi $\mathcal{G}_s(\mathcal{N}_s, \mathcal{L}_s)$ on graafin $\mathcal{G}(\mathcal{N}, \mathcal{L})$ aligraafi (*subgraph*), jos

$$\mathcal{N}_s \subseteq \mathcal{N} \text{ ja } \mathcal{L}_s \subseteq \mathcal{L}.$$

Graafin $\mathcal{G}(\mathcal{N}, \mathcal{L})$ solmugeneroitu aligraafi (*node-generated subgraph*) on graafi $\mathcal{G}_s(\mathcal{N}_s, \mathcal{L}_s)$ se., kaarijoukko \mathcal{L}_s sisältää kaikki ne kaarijoukon \mathcal{L} kaaret, jotka yhdistävät solmuja joukossa \mathcal{N}_s . Graafin $\mathcal{G}(\mathcal{N}, \mathcal{L})$ viivageneroitu aligraafi (*line-generated subgraph*) on graafi $\mathcal{G}_s(\mathcal{N}_s, \mathcal{L}_s)$ se., solmujoukko \mathcal{N}_s sisältää kaikki ne solmujoukon \mathcal{N} solmut, jotka ovat liittyneitä joukon \mathcal{L}_s kaariin.

Dyadi (*a dyad, dyads*) on solmugeneroitu aligraafi, joka koostuu solmuparista ja mahdollisesta solmuja yhdistävästä kaaresta. Triadi (*a triad, triads*) on myöskin solmugeneroitu aligraafi, jossa solmuja on kolme ja solmuja yhdistäviä kaaria 0-3 kappaletta (ks. kuva 2.3).



Kuva 2.3: Triadissa on kaaria 0-3 kappaletta.

Digraafien aligraafit määritellään samalla tavalla kuin on esitetty edellä suuntaamattomille graafeille. Digraafeissa dyadeille määritellään kuitenkin nk. isomorfismiluokkia (*isomorphism classes*), joita on syytä käsitellä tarkemmin.

Dyadi on yksi tärkeimmistä digraafin aligraafeista. Digraafin dyadin solmuja n_i ja n_j voi yhdistää jompaan kumpaan suuntaan osoittava nuoli tai vastakkaisiin suuntiin osoittavat kaksi nuolta. Toisaalta kaksi toisiinsa kytkeytymätöntä solmua muodostavat myös dyadin. Täten digraafissa dyadi voi olla neljässä eri tilassa, mutta nämä tilat voidaan jakaa kolmeen isomorfismiluokkaan.

Nolla-dyadissa (*a null dyad*), solmuparin välillä ei ole nuolta kumpaankaan suuntaan. Ts. solmujen n_i ja n_j muodostama dyadi on nolla-dyadi, jos

$$\langle n_i, n_j \rangle \notin \mathcal{L} \quad \wedge \quad \langle n_j, n_i \rangle \notin \mathcal{L}.$$

Asymmetrinen dyadi (*an asymmetric dyad*), muodostuu solmuparista ja niiden yhdistävästä ainoastaan toiseen suuntaan osoittavasta nuolesta. Ts. solmujen n_i ja n_j muodostama dyadi on asymmetrinen, jos ja vain jos

$$\langle n_i, n_j \rangle \notin \mathcal{L} \quad \oplus \quad \langle n_j, n_i \rangle \notin \mathcal{L},$$

missä \oplus tarkoittaa poissulkevaa tai-konnektiivia (*XOR, exclusive or*), jonka totuusarvot on esitetty taulukossa 2.1. Logiikassa poissulkeva tai-konnektiivi palauttaa saman totuusarvon kuin tai-konnektiivi aina, paitsi silloin kuin lauseista p ja q molemmat ovat totta. Tällöin poissulkeva tai palauttaa totuusarvon epätosi.

Taulukko 2.1: OR ja XOR.

p	q	\vee	\oplus
0	0	0	0
0	1	1	1
1	0	1	1
1	1	1	0

Yksi solmupari voi muodostaa kaksi erilaista asymmetristä dyadia, riippuen solmuja yhdistävän nuolen suunnasta. Nämä kaksi erilaista dyadia ovat kuitenkin isomorfisia (ks. luku 2.2.7) keskenään.

Kaksi solmua n_i ja n_j muodostavat keskinäisen dyadin (*mutual, reciprocal dyad*), kun solmujen välillä on yhteys molempiin suuntiin, ts.

$$\langle n_i, n_j \rangle \in \mathcal{L} \quad \wedge \quad \langle n_j, n_i \rangle \in \mathcal{L}.$$

Keskinäisessä dyadissa solmuja yhdistää siis kaksi nuolta, jotka osoittavat vastakkaisiin suuntiin solmujen välillä.

2.2.3 Solmun aste

Määritelmä 2.2.2. Suuntaamattomassa graafissa solmun aste (*degree*) $d(n_i)$ on yhtä suuri kuin solmuun n_i liittyneiden kaarien lukumäärä.

Kun graafissa \mathcal{G} on g solmua, kullakin solmulla n_i on aste, joka voi olla minimissään 0, jolloin solmuun n_i ei ole liittynyt yhtään kaarta, ja maksimissaan $g-1$, jolloin kaikki muut graafin solmut ovat liittyneet suoraan erillisillä kaarilla solmuun n_i .

Määritelmä 2.2.3. Graafin \mathcal{G} , jossa on g solmua ja L kaarta, solmujen keskiaste (*mean degree*) määritellään

$$\bar{d} = \frac{\sum_{i=1}^g d(n_i)}{g} = \frac{2L}{g}.$$

Edelleen astelukujen varianssi (*variance of degrees*) määritellään

$$S_d^2 = \frac{\sum_{i=1}^g (d(n_i) - \bar{d})^2}{g} \quad (2.2.4)$$

Määritelmä 2.2.5. Jos graafin jokaisen solmun aste on yhtä suuri, eli

$$d(n_i) = d, \forall i \in \{1, 2, \dots, g\}, d \in \{0, 1, \dots, g-1\}, g \in \{0, \mathbb{N}\},$$

niin graafi on säännöllinen (*regular*).

Tällaisen d -säännöllisen (*d-regular*) graafin astelukujen varianssi $S_d^2 = 0$.

Suunnatuilla graafeilla solmuille määritellään kaksi toisistaan poikkeavaa astelukua. Suunnattu yhteys kulkee solmuparin välillä se, toinen solmuista on yhteyden alkusolmu ja toinen vastaanottajasolmu. Näin ollen digraafin solmujen astelukuja tarkasteltaessa tulee huomioida solmuun kohdistuneiden yhteyksien lukumäärä ja solmusta lähtevien yhteyksien lukumäärä.

Määritelmä 2.2.6. Digraafissa solmun tuontiluku (*indegree*) $d_I(n_i)$ on yhtä suuri kuin solmuun n_i tulevien/osoittavien nuolien lukumäärä.

Määritelmä 2.2.7. Digraafissa solmun vientiluku (*outdegree*) $d_O(n_i)$ on yhtä suuri kuin solmusta n_i lähtevien nuolien lukumäärä.

Kun digraafissa \mathcal{G}_d on g solmua, kullakin solmulla n_i on tuontiluku, joka voi olla minimissään 0, jolloin solmuun n_i ei osoita yhtään nuolta muista solmuista, ja maksimissaan $g-1$, jolloin kaikista graafin muista solmuista osoittaa nuoli solmuun n_i . Vastaavasti kullakin solmulla n_i on vientiluku, joka on minimissään 0, kun solmusta n_i ei lähde yhtään nuolta muihin solmuihin, ja maksimissaan $g-1$, kun solmusta n_i osoittaa erillinen nuoli digraafin jokaiseen muuhun solmuun.

Määritelmä 2.2.8. Digraafin \mathcal{G}_d , jossa on g solmua ja L kaarta, solmujen keskituontiluku (*mean indegree*) määritellään

$$\bar{d}_I = \frac{\sum_{i=1}^g d_I(n_i)}{g} = \frac{L}{g}.$$

Määritelmä 2.2.9. Digraafin \mathcal{G}_d , jossa on g solmua ja L kaarta, solmujen keskivientiluku (*mean outdegree*) määritellään

$$\bar{d}_O = \frac{\sum_{i=1}^g d_O(n_i)}{g} = \frac{L}{g}.$$

Lause 2.2.10. Digraafin \mathcal{G}_d , jossa on g solmua ja L kaarta, solmujen keskivientiluvulle ja keskituontiluvulle (*mean indegree*) pätee

$$\bar{d}_O = \bar{d}_I = \frac{L}{g}.$$

Todistus. Koska tuontilukuja summattaessa lasketaan yhteen digraafin kaikkien nuolien kärjet, vastaa tämä summa $\sum_{i=1}^g d_I(n_i)$ digraafin nuolien lukumäärää L . Vastaavasti vientilukuja summattaessa lasketaan yhteen digraafin kaikkien nuolten "alkupisteet", milloin summa $\sum_{i=1}^g d_O(n_i)$ vastaa myöskin digraafin nuolien lukumäärää L . Täten siis

$$\bar{d}_I = \bar{d}_O = \frac{L}{g}.$$

□

Tuonti- ja vientiluvuille määritellään myös varianssit. Tuontilukujen varianssi määritellään

$$S_{d_I}^2 = \frac{\sum_{i=1}^g (d_I(n_i) - \bar{d}_I)^2}{g}, \quad (2.2.11)$$

ja vastaavasti vientiluvuille

$$S_{d_O}^2 = \frac{\sum_{i=1}^g (d_O(n_i) - \bar{d}_O)^2}{g}. \quad (2.2.12)$$

Molempia variansseja käytetään tunnistamaan, kuinka erilaisia verkoston toimijat ovat suhteessa yhteyksien vastaanottamiseen ja niiden lähettämiseen. Tällaisen statistiikan avulla voidaan yksinkertaisesti sanoa kuinka keskittynyt (*centralized*) verkosto on. Verkoston keskittyneisyyden tunnuslukuja käsitellään luvussa 2.6.

2.2.4 Graafin tiheys

Määritelmä 2.2.13. Graafin $\mathcal{G}(\mathcal{N}, \mathcal{L})$ tiheys (*density*) Δ on graafin kaarien osuus graafin kaikista mahdollisista kaarista.

Lause 2.2.14. Graafin $\mathcal{G}(\mathcal{N}, \mathcal{L})$, jossa on L kaarta ja g solmua, tiheydelle Δ pätee

$$\Delta = \frac{2L}{g(g-1)}.$$

Todistus. Yksinkertaisessa (ei silmukoita, eikä rinnakkaisia (*parallel*) kaaria) graafissa, jossa on g solmua, on kaaria maksimissaan

$$\binom{g}{2} = \frac{g!}{2!(g-2)!} = \frac{g(g-1)(g-2)!}{2!(g-2)!} = \frac{g(g-1)}{2}.$$

Koska tiheys määritellään olemassa olevien kaarien ja graafin maksimaalisen kaarien määrän suhteena, on tällöin

$$\Delta = \frac{L}{g(g-1)/2} = \frac{2L}{g(g-1)}.$$

□

Tiheys on minimissään 0, jos graafissa ei ole lainkaan kaaria ($L = 0$) ja maksimissaan 1, jos graafi on täydellinen (*complete*) eli graafin kaikki mahdolliset kaaret ovat edustettuina ($L = g(g-1)/2$). Täydellistä graafia, jossa on g solmua, merkitään K_g .

Suuntaamattomassa g solmun graafissa kaikkien solmujen asteiden summa on $2L$, mikä antaa keskiasteeksi $2L/g$. Täten tiheys voidaan kirjoittaa muodossa

$$\Delta = \frac{2L}{g(g-1)} = \frac{\bar{d}}{g-1}. \quad (2.2.15)$$

Tämä tiheyden määritelmä pätee myös aligraafeille $\mathcal{G}_s(\mathcal{N}_s, \mathcal{L}_s)$

$$\Delta = \frac{2L_s}{g_s(g_s-1)}. \quad (2.2.16)$$

Digraafin tiheys määritellään vastaavalla tavalla digraafissa olevien nuolien määrän suhteena digraafin maksimaaliseen nuolten määrään.

Lause 2.2.17. Digraafin $\mathcal{G}_d(\mathcal{N}, \mathcal{L})$, jossa on L nuolta ja g solmua, tiheydelle Δ pätee

$$\Delta = \frac{L}{g(g-1)}.$$

Todistus. Koska digraafissa jokaisesta solmusta voi osoittaa nuoli yhteensä $(g-1)$:een solmuun, on maksimaalinen nuolien määrä siis $g(g-1)$. Tiheys määritellään olemassa olevien nuolten määrän suhteena maksimaaliseen nuolten määrään, joten

$$\Delta = \frac{L}{g(g-1)}.$$

□

2.2.5 Kulku, reitti ja polku

Määritelmä 2.2.18. Kulku (*a walk, walks*) on graafin $\mathcal{G}(\mathcal{N}, \mathcal{L})$ solmujen ja kaarien äärellinen jakso

$$W = n_{i0}, l_{j0}, n_{i1}, l_{j1}, \dots, l_{jk}, n_{ik}.$$

Kulku alkaa aina solmusta ja päättyy solmuun. Mikäli kulun alkusolmu n_{i0} ja loppusolmu n_{ik} ovat yksi ja sama solmu n , on kulku suljettu (*closed*). Kulku voi sisältää saman kaaren ja solmun useammin kuin kerran jaksossa. Kulun pituus (*length*) on kulun sisältämien kaarien lukumäärä. Jos jokin kaari on kulussa useampaan kertaan, lasketaan ne erillisinä kaarina kulun pituuteen.

Kulun vastakulku (*inverse*) W^{-1} on itse kulku käänteisessä järjestyksessä. Ts.,

$$W^{-1} = n_{ik}, l_{jk}, n_{i(k-1)}, l_{j(k-1)}, \dots, n_{i1}, l_{j1}, l_{j0}, n_{i0}.$$

Reitti (*a trail, trails*) on kulku, jossa kukin kaari esiintyy vain kerran. Polku (*a path, paths*) puolestaan on kulku, jossa kukin solmu ja kaari esiintyy vain kerran. Piiri (*a cycle, cycles*) on suljettu kulku, jossa on vähintään kolme solmua, joista kaikki

muut solmut paitsi alkuperä- ja loppusolmu ovat erillisiä sekä kaikki kulun sisältämät kaaret ovat erillisiä. Hamiltonin piiri (*Hamiltonian cycle*) kulkuun sisältyy graafin jokainen kaari ja solmu ainoastaan kerran (pl. alkuperä- ja loppusolmu, joka on yksi ja sama solmu). Kierros (*a tour, tours*) on suljettu kulku, jossa jokainen graafin kaari on kuljettu vähintään kerran.

Digraafeille määritellään kulkua, reittiä ja polkua vastaavat käsitteet. Erona graafeihin on ainoastaan se, että digraafeissa tulee ottaa huomioon yhteyksien suunnat.

Suunnattu kulku (*a directed walk*) on digraafin solmujen ja nuolien äärellinen jakso, niin että jokaisen jaksossa olevan nuolen alkusolmu on nuolta edeltävä solmu. Voidaan siis ajatella, että suunnatun kulun kaikki nuolet osoittavat "samaa suuntaa". Suunnatun kulun pituus lasketaan sen sisältämien nuolien lukumääränä. Suunnattu kulku voi sisältää saman nuolen tai solmun useammin kuin kerran jaksossa, ja jos jokin kaari on suunnatussa kulussa edustettuna useampaan kertaan, lasketaan ne erillisinä kulun pituuteen.

Suunnattu reitti (*a directed trail*) on digraafin solmujen ja nuolien äärellinen jakso, jossa kukin nuoli esiintyy korkeintaan kerran. Edelleen suunnatussa polussa (*a directed path*) jaksoon kuuluva solmu tai nuoli esiintyy ainoastaan kerran.

Digraafien kuluissa ja poluissa ei kuitenkaan ole aina välttämätöntä huomioida ajatusta nuolten "samasta suunnasta", vaan eteneminen nuolen suuntaa vastaan voidaan myös sallia. Tällöin puhutaan puolikulusta (*a semiwalk*) ja -polusta (*a semipath*). Näiden määritelmät vastaavat suunnatun kulun ja suunnatun polun yleisiä määritelmiä, mutta yhteyden suunnalla ei ole merkitystä.

Digraafissa piiri on suljettu suunnattu kulku, jossa on vähintään kolme solmua, joista kaikki muut solmut paitsi alkuperä- ja loppusolmu ovat erillisiä sekä kaikki kulun sisältämät nuolet ovat erillisiä. Mikäli piiri muodostaa suljettu suunnattu kulku on puolikulku, sanotaan piiriä puolipiiriksi (*a semicycle*).

2.2.6 Graafin yhtenäisyys

Määritelmä 2.2.19. Graafi on yhtenäinen (*connected*), jos sen jokaisen solmuparin välillä on polku.

Yhtenäisessä graafissa jokainen solmu on saavutettavissa (*reachable*) toisesta solmusta. Jos graafi ei ole yhtenäinen, se on epäyhtenäinen (*disconnected*). Epäyhtenäisen graafin maksimaalisia yhtenäisiä aligraafeja sanotaan graafin komponenteiksi (*a component, components*). Maksimaalisella tarkoitetaan tässä sitä, että aligraafin ei voida lisätä ainuttakaan graafin muuta solmua, joka ei vielä kuulu aligraafin solmujoukkoon se., solmun lisäyksen jälkeen uudessa aligraafissa jokaisen solmuparin välillä olisi polku.

Digraafin yhtenäisyys riippuu solmuja yhdistävien nuolien suunnasta ja siitä,

ovatko solmuparit yhteydessä toisiinsa polun vai puolipolun kautta. Solmut n_i ja n_j voivat olla yhteydessä toisiinsa polun tai puolipolun kautta neljällä eri tavalla. Nämä eri tapaukset määrittelevät solmuparin yhteyden tyyppin, joka voi olla:

1. Heikosti yhdistetty (*weakly connected*), jos solmuparia yhdistää puolipolku
2. Yksipuolisesti yhdistetty (*unilaterally connected*), jos solmuparia yhdistää polku solmusta n_i solmuun n_j tai päinvastoin
3. Vahvasti yhdistetty (*strongly connected*), jos solmuparia yhdistää sekä polku solmusta n_i solmuun n_j että polku solmusta n_j solmuun n_i , mutta nämä polut eivät välttämättä sisällä samoja solmuja ja nuolia
4. Rekursiivisesti yhdistetty (*recursively connected*), jos solmupari on vahvasti yhdistetty ja polku solmusta n_i solmuun n_j ja polku solmusta n_j solmuun n_i sisältävät samat solmut ja nuolet, mutta vastakkaisessa järjestyksessä

Täten riippuen digraafin solmuparien yhtenäisyyksistä digraafin yhtenäisyys voi olla jotakin seuraavista neljästä tyypestä:

1. Heikosti yhtenäinen (*weakly connected*), jos kaikki digraafin solmuparit ovat heikosti yhdistettyjä
2. Yksipuolisesti yhtenäinen (*unilaterally connected*), jos kaikki digraafin solmuparit ovat yksipuolisesti yhdistettyjä
3. Vahvasti yhtenäinen (*strongly connected*), jos kaikki digraafin solmuparit ovat vahvasti yhdistettyjä
4. Rekursiivisesti yhtenäinen (*recursively connected*), jos kaikki digraafin solmuparit ovat rekursiivisesti yhdistettyjä

Etäisyys ja geodeesi

Määritelmä 2.2.20. Geodeesi (*a geodesic, geodesics*) on lyhin polku graafin kahden solmun välillä.

Geodeettinen etäisyys (*geodesic distance*), tai yksinkertaisemmin etäisyys (*distance*), kahden solmun välillä määritellään solmujen geodeesin pituutena.

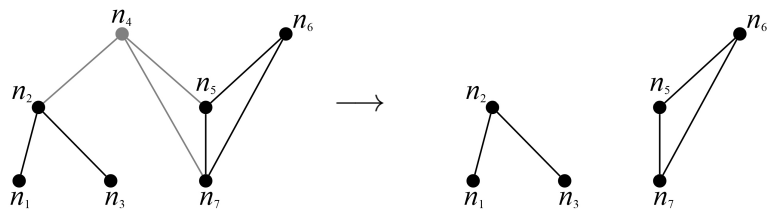
Etäisyyttä solmujen n_i ja n_j välillä merkitään $d(i,j)$. Solmujen n_i ja n_j välinen etäisyys on minkä tahansa geodeesin pituus ko. solmujen välillä. Mikäli solmuparin välillä ei ole polkua, on ko. solmujen välinen etäisyys ääretön (tai määrittämätön). Suuntaamattomilla graafeilla $d(i,j) = d(j,i)$.

Graafin halkaisija (*diameter*) on yhtä suuri kuin graafin kaikkien geodeettisten etäisyyksien maksimi. Solmun n_i eksentrisyys (*eccentricity*) eli epäkeskisyys (myös ns. suhdeluku (*association number*)) on suurin geodeettinen etäisyys solmun n_i ja minkä tahansa graafin muun solmun n_j kanssa.

Digraafeissa geodeesi määritellään vastaavalla tavalla, kuin graafeissa. Huomatavaa on kuitenkin, ettei digraafeilla aina päde $d(i,j) = d(j,i)$. Edelleen digraafin halkaisija määritellään kuten graafin halkaisija. Jotta digraafin halkaisija olisi määritettävissä, tulee digraafin olla joko vahvasti tai rekursiivisesti yhtenäinen (ts. jokaisen solmuparin välillä on jokin suunnattu polku). Jos digraafi on kuitenkin vain yksipuolisesti tai heikosti yhtenäinen, on tällöin jotkin digraafin geodeesit äärettömiä (tai määrittämättömiä). Yksipuolisesti ja heikosti yhtenäisen digraafin halkaisija on täten määrittämätön.

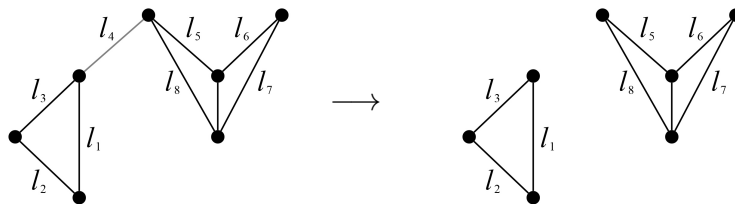
Piste- ja viivayhtenäisyys

Graafin solmu n_i on irrotuspiste (*a cutpoint, cutpoints*), jos graafissa, joka sisältää $n_i:n$, on vähemmän komponentteja kuin graafissa, josta n_i ja siihen liittyneet kaaret on poistettu.



Kuva 2.4: Solmu n_4 on irrotuspiste.

Graafin kaari l_i on silta (*a bridge, bridges*), jos graafissa, joka sisältää $l_i:n$, on vähemmän komponentteja kuin graafissa, josta kaari l_i on poistettu.



Kuva 2.5: Kaari l_4 on silta.

Solmupisteen ja sillan määritelmät ovat suoraan päteviä myös digraafeille.

Graafin \mathcal{G} piste- tai solmuyhtenäisyys (*point-/node-connectivity*) $\kappa(\mathcal{G})$ on pienin luku κ , jolla graafissa on κ -pisteirrotus. κ -pisteirrotuksella tarkoitetaan pienintä graafin \mathcal{G} solmujen lukumäärää, mikä tulisi poistaa graafista, jotta saataisiin epäyhtenäinen graafi tai triviaali graafi.

- Jos graafi on valmiiksi epäyhtenäinen $\kappa = 0$
- Jos graafissa on irrotuspiste n_i , niin $\kappa = 1$
- Jos graafista pitää poistaa n solmua, jotta graafi olisi epäyhtenäinen, niin $\kappa = n$
- Jos graafi on täydellinen K_g -graafi, niin $\kappa = g - 1$
- Mille tahansa luvulle $k < \kappa$ graafin sanotaan olevan k -solmuyhtenäinen (*k-node connected*)

Graafin \mathcal{G} viiva- tai kaariyhtenäisyys (*line-/edge-connectivity*) $\lambda(\mathcal{G})$ on pienin luku λ , jolla graafissa on λ -viivairrotus. λ -viivairrotuksella tarkoitetaan pienintä graafin \mathcal{G} kaarien lukumäärää, mikä tulisi poistaa graafista, jotta saataisiin epäyhtenäinen graafi tai triviaali graafi.

- Jos graafi on valmiiksi epäyhtenäinen $\lambda = 0$
- Jos graafissa on silta l_i , niin $\lambda = 1$
- Jos graafista pitää poistaa m kaarta, jotta graafi olisi epäyhtenäinen, niin $\lambda = m$
- Jos graafi on täydellinen K_g -graafi, niin $\lambda = g - 1$
- Mille tahansa luvulle $l < \lambda$ graafin sanotaan olevan l -viivayhtenäinen (*l-line connected*)

2.2.7 Isomorfiset graafit

Määritelmä 2.2.21. Graafit $\mathcal{G}(\mathcal{N}, \mathcal{L})$ ja $\mathcal{G}^*(\mathcal{N}^*, \mathcal{L}^*)$ ovat isomorfisia, jos on olemassa bijektio $\phi : \mathcal{N} \rightarrow \mathcal{N}^*$ se.,

$$\begin{aligned} \exists \phi(n_i) = n_k^*, \quad \exists \phi(n_j) = n_l^* : \forall n_i, n_j \in \mathcal{N}, \quad \forall n_k^*, n_l^* \in \mathcal{N}^* \\ l_m = (n_i, n_j) \in \mathcal{L} \iff l_o = (n_k^*, n_l^*) \in \mathcal{L}^*. \end{aligned}$$

Kaksi graafia \mathcal{G} ja \mathcal{G}^* ovat isomorfisia (*isomorphic*), jos on olemassa yksi yhteen kuvaus \mathcal{G} :n solmuilta \mathcal{G}^* :n solmuille se., solmujen vierekkäisyydet säilyvät. Yksi yhteen kuvauksella tarkoitetaan, että jokainen \mathcal{G} :n solmu kuvautuu yhteen (ja vain yhteen) \mathcal{G}^* :n solmuun ja päinvastoin. Toisin sanoen, jos graafien $\mathcal{G}(\mathcal{N}, \mathcal{L})$ ja $\mathcal{G}^*(\mathcal{N}^*, \mathcal{L}^*)$ solmujoukkojen välillä on yksi yhteen kuvaus ϕ , niin tällöin $\phi(n_i) = n_k^*$, ja sen käänteiskuvaus ϕ^{-1} toteuttaa ehdon $\phi^{-1}(n_k^*) = n_i$.

2.2.8 Graafin komplementti

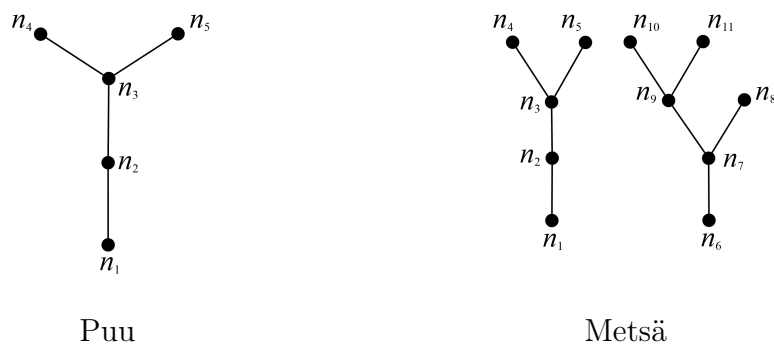
Graafin \mathcal{G} , jossa on g solmua, komplementissa (*complement*) $\overline{\mathcal{G}}$ on sama solmujoukko kuin \mathcal{G} :ssä, mutta sen kaarijoukkoon kuuluvat ainoastaan ne kaaret, jotka eivät esiinny \mathcal{G} :ssä, mutta esiintyisivät täydellisessä K_g -graafissa. Ts. jos solmut n_i ja n_j ovat vierekkäiset \mathcal{G} :ssä, niin ne eivät ole vierekkäiset $\overline{\mathcal{G}}$:ssä ja päinvastoin.

Digraafin \mathcal{G}_d komplementissa $\overline{\mathcal{G}_d}$ on sama solmujoukko kuin \mathcal{G}_d :ssä, mutta sen nuolijoukkoon kuuluvat ainoastaan ne nuolet, jotka eivät esiinny \mathcal{G}_d :ssä. Ts., jos nuoli $\langle n_i, n_j \rangle$ on \mathcal{G}_d :ssä, niin se ei ole $\overline{\mathcal{G}_d}$:ssä, jos taas nuoli $\langle n_i, n_j \rangle$ ei ole \mathcal{G}_d :ssä, niin se on $\overline{\mathcal{G}_d}$:ssä.

Digraafeille määritellään myös vastakohta (*converse*), joka poikkeaa komplementista oleellisesti. Digraafin vastakohdassa on sama solmujoukko kuin alkuperäisessä digraafissa, mutta sen nuolijoukkoon kuuluvat nuolet, jotka ovat vastakkaissuuntaisia alkupeäisen digraafin nuolien kanssa. Vastakohta saadaan siis alkuperäisestä digraafista kääntämällä sen kaikkien nuolien suunta päinvastaiseksi. Mikäli alkupeäisessä digraafissa on läsnä solmuparia yhdistäviä vastakkaisia nuolipareja, säilyvät nämä nuoliparit sellaisenaan digraafin vastakohdassa.

2.2.9 Puu ja metsä

Puu (*a tree, trees*) on asyklinen (*acyclic*) eli piiritön yhtenäinen (di)graafi. Piiritön epäyhtenäinen (di)graafi on metsä (*a forest, forests*), jos sen jokainen (di)graafin komponentti on puu. Kuvassa 2.6 on esitetty erikseen suuntaamaton puu ja suuntaamaton metsä.

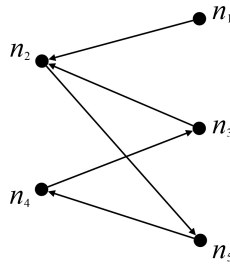


Kuva 2.6: Puu ja metsä.

2.2.10 Kaksijakoinen graafi

Määritelmä 2.2.22. Kaksijakoisessa (*bipartite*) (di)graafissa solmujoukko \mathcal{N} voidaan jakaa kahteen alisolmujoukkoon \mathcal{N}_1 ja \mathcal{N}_2 se., jokainen joukon \mathcal{L} kaari l_k on järjestetty pari se.,

$$l_k = \langle n_i, n_j \rangle, \quad n_i \in \mathcal{N}_1, \quad n_j \in \mathcal{N}_2.$$



Kuva 2.7: Kaksijakoinen digraafi.

Täydellisessä kaksijakoisessa graafissa (*complete bipartite graph*) jokainen joukon \mathcal{N}_1 on vierekkäinen jokaisen joukon \mathcal{N}_2 solmun kanssa. Täydellistä kaksijakoista graafia, jonka solmujoukossa \mathcal{N}_1 on yhteensä g_1 solmua ja solmujoukossa \mathcal{N}_2 yhteensä g_2 solmua, merkitään K_{g_1, g_2} .

Kaksijakoista graafia (ks. kuva 2.7) voidaan tietyissä tapauksissa käyttää yhtenä esitystapana kaksimoodiselle verkostolle. Tästä puhutaan lisää luvussa 2.7.2.

2.3 Matriisit SNA:ssa

Verkoston analysoinnissa SNA-menetelmät pohjautuvat pääosin matriisilaskennan menetelmiin. Tarkasteltava verkosto voidaan mallintaa matriisina, jonka avulla laskennallinen analyysi on mahdollista suorittaa tehokkaasti suurillekin verkostoille. Matriisien avulla voidaan helposti esittää tietoja verkostosta, verkoston toimijoista ja niiden välisistä suhteista.

Matriisien käyttö SNA:ssa vaatii tarkempia määritelmiä tarkastettaville verkostoille, mutta samalla mahdollistaa monipuolisemman analyysin. Verkoston matriisimalliin voidaan tallentaa tietoa toimijoiden välisten suhteiden olemassaolosta sekä tarvittaessa suhteiden voimakkuudesta, jos ne pystytään määrittämään numeerisina lukuarvoina. Täten SNA:ssa käytetyt matriisit tuovat mukanaan toisenlaisen kahtiajaon tarkasteltaville verkostoille. Mikäli verkostossa toimijoiden välisissä yhteyksissä

otetaan huomioon niiden voimakkuus, sanotaan verkoston yhteyksien olevan arvoitettuja (*valued*), ja jos taas yhteyksiä tarkasteltaessa kiinnostaa vain niiden olemassaolo, puhutaan arvottomattomista (*non-valued*) tai dikotomisista (*dichotomous*) yhteyksistä.

SNA:ssa verkoston matriisimalleja on useita. Eri matriisit soveltuvat eri tavoin erilaisten verkostojen mallintamiseen ja analyysiin. Tärkeintä on, että matriisien avulla tehtävässä analyysissä käytetään verkostoa parhaiten mallintavaa matriisiä, jotta laskennallisessa analyysissä voidaan ottaa mahdollisimman monipuolisesti verkostoa koskevat tiedot huomioon.

2.3.1 Sosiomatriisi

Verkoston toimijat, jotka graafeissa esitetään solmuina, ja eri toimijaparien väliset (suunnatut tai suuntaamattomat) yhteydet voidaan esittää yhtenä matriisina. Jos verkosto koostuu g :stä toimijasta, joiden välillä joko on yhteys tai yhteys puuttuu, voidaan toimijoiden väliset yhteydet kuvata taulukkona (ks. taulukko 2.2), jossa kullekin toimijalle on merkitty oma vaakarivi ja vastaava pystysarake. Taulukkoon merkitään binääriluvuilla 0 ja 1 toimijoiden välisen yhteyden olemassa olo se., alkio saa arvon 0, jos yhteyttä ei ole, ja arvon 1, jos toimijoiden välillä on yhteys. Koska silmukoita ei sallita verkostossa, taulukon lävistäjän alkio jätetään määrittelemättä.

Taulukko 2.2: Sosiaalisen verkoston taulukkoesitys.

	n_1	n_2	n_3	n_4
n_1	—	1	0	1
n_2	1	—	0	0
n_3	0	0	—	1
n_4	1	0	1	—

Diagonaalit kuvaavat yhteyttä toimijaan itseensä, ja ne voidaan joko jättää määrittelemättä tai merkitä nolliksi, kun käsitellään yksinkertaisia, silmukattomia, graafeja.

Taulukko on $g \times g$ vieruspistematriisi (*an adjacency matrix*) (Ruohonen, 2006, 32) \mathbf{X} .

Määritelmä 2.3.1. Vieruspistematriisin \mathbf{X} alkio x_{ij} määritellään

$$x_{ij} = \begin{cases} 0, & \text{jos } l_k = (n_i, n_j) \text{ ei ole olemassa} \\ 1, & \text{jos } l_k = (n_i, n_j) \text{ on olemassa.} \end{cases}$$

Tätä vieruspistematriisiä nimitetään SNA:ssa sosiomatriisiksi (*a sociomatrix, sociomatrixes*). Edellä esitetystä taulukosta 2.2 saadaan siis sosiomatriisi \mathbf{X}

$$\mathbf{X} = \begin{bmatrix} - & 1 & 0 & 1 \\ 1 & - & 1 & 0 \\ 0 & 1 & - & 1 \\ 0 & 1 & 0 & - \end{bmatrix}$$

Sosiomatriisi on yleisesti asymmetrinen (*asymmetric*) suunnatuille graafeille, mutta aina symmetrinen (*symmetric*) suuntaamattomille graafeille. Täydellisen K_g -graafin sosiomatriisin jokainen diagonaalialkiosta poikkeava alkio on arvoltaan 1. Vastaavasti tyhjää graafia vastaa sosiomatriisi, jossa jokainen (diagonaalialkiosta poikkeava) alkio on arvoltaan 0. Arvotetuille graafeille sosiomatriisin alkiot ovat reaalityyppisiä, jotka vastaavat toimijoiden välisten yhteyksien arvoja v_k . Tästä puhutaan lisää luvussa 2.5.

2.3.2 Insidenssimatriisi

Insidenssimatriisissa (*an incidence matrix*) on esitetty tieto siitä, mitkä graafin (verkoston) solmut ovat johtuvia (*incident*) minkäkin graafin (verkoston) kaaren suhteen. Insidenssimatriisissa kutakin solmua vastaa yksi matriisin rivi ja kutakin kaarta yksi sarake.

Määritelmä 2.3.2. Verkoston, jossa on g solmua ja L kaarta, insidenssimatriisi \mathbf{I} on $g \times L$ matriisi, jonka alkiot I_{ij} ovat binäärisiä se.,

$$I_{ij} = \begin{cases} 0, & \text{jos solmu } n_i \text{ ei ole liittynyt kaareen } l_j \\ 1, & \text{jos solmu } n_i \text{ on liittynyt kaareen } l_j. \end{cases}$$

Insidenssimatriisin jokaisessa sarakkeessa on täsmälleen kaksi ykköstä niillä riveillä, jotka edustavat kyseisen kaaren päätepisteitä. Insidenssimatriisi saadaan vastaavalla tavalla omasta taulukkoesityksestään, kuten sosiomatriisi.

	l_1	l_2	l_3	l_4	l_5		
n_1	1	1	0	0	1	\longrightarrow	$\mathbf{I} =$
n_2	1	0	1	0	0		
n_3	0	1	1	1	0		
n_4	0	0	0	1	1		

$$\mathbf{I} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

2.3.3 Muut SNA-matriisit

Kirjallisuudessa esiintyy myös vieruspiste- ja insidenssimatriisien lisäksi muita SNA-matriiseja. Niiden käyttö on kuitenkin rajattua ja vaatii tietynlaisen tulkinnan verkoston ominaisuuksista ja rakenteesta (toimijatyypit, yhteydet, jne.)

Kaksimoodisen verkoston (*two-mode network*) sosiomatriisi on $g \times h$ matriisi, jossa kutakin solmua joukosta $\mathcal{N} = \{n_1, n_2, \dots, n_g\}$ vastaa yksi rivi ja kutakin

solmua joukosta $\mathcal{M} = \{m_1, m_2, \dots, m_h\}$ vastaa yksi sarake. Kaksimoodisen verkoston matriiseja on käsitelty lähemmin luvussa 2.7.2.

Hypergraafia (*hypergraph*) $\mathcal{H}(\mathcal{N}, \mathcal{M})$ kuvaava matriisi \mathbf{A} on $g \times h$ matriisi, jossa alkio $a_{ij} = 1$, jos pistejoukon $\mathcal{N} = \{n_1, n_2, \dots, n_g\}$ piste n_i on kaarijoukon $\mathcal{M} = \{M_1, M_2, \dots, M_h\}$ kaaressa M_j , ja 0 muulloin. Matriisia \mathbf{A} on sanottu hypergraafin insidenssmatriisiksi, koska se ilmoittaa mitkä pisteet ovat liittyneet mihinkin kaariin.

2.4 SNA-matriisien ominaisuuksia

Tässä luvussa käsitellään pääosin sosiomatriisin ominaisuuksia laskennallisessa SNA:ssa. Esiteltävät ominaisuudet ovat pohjana syvemmälle SNA:lle, jolla saadaan tietoa mm. verkoston rakenteesta ja toimijoista. Näitä kuvaavat verkoston tunnusluvut, joita käsitellään luvussa 2.6.

2.4.1 Kulku

Sosiomatriisin \mathbf{X} alkiot x_{ij} kertovat, onko solmujen n_i ja n_j välillä kulku $n_i n_j$. Sosiomatriisin \mathbf{X} neliön \mathbf{X}^2 alkio $x_{ij}^{[2]}$ määritellään

$$x_{ij}^{[2]} = \sum_{k=1}^g x_{ik} x_{kj}. \quad (2.4.1)$$

Tämän summan yksi termi $x_{ik} x_{kj} = 1$ vain, jos molemmat yhteydet (n_i, n_k) ja (n_k, n_j) ovat olemassa. Summa laskee siis kulkujen, joiden pituus on kaksi, lukumäärän solmusta n_i solmuun n_j .

Sosiomatriisin \mathbf{X} neliön \mathbf{X}^2 alkiot $x_{ij}^{[2]}$ ilmoittavat verkostossa olevien kulkujen, joiden pituus on kaksi, lukumäärän solmusta n_i solmuun n_j . Edelleen matriisin \mathbf{X}^p alkiot ilmoittavat solmujen välisien kulkujen, joiden pituus on p , lukumäärän.

2.4.2 Saavutettavuus

Saavutettavuudella (*reachability*) tarkoitetaan sitä, onko joidenkin verkoston kahden solmun välillä kulku. Saavutettavuusmatriiseissa (*a reachability matrix*) $\mathbf{X}^{[R]}$ alkio $x_{ij}^{[R]}$ on yksi, jos solmujen n_i ja n_j välillä on kulku, ja nolla muulloin.

Koska g -solmuisessa verkostossa kulku voi olla pituudeltaan korkeintaan $g-1$, sosiomatriisin \mathbf{X} potenssit $\mathbf{X}^2, \mathbf{X}^3, \dots, \mathbf{X}^{g-1}$ ilmoittavat solmujen välisten erimittaiten kulkujen lukumäärät. Näiden summamatriisi

$$\mathbf{X}^{[\Sigma]} = \sum_{i=1}^{g-1} \mathbf{X}^i = \mathbf{X} + \mathbf{X}^2 + \mathbf{X}^3 + \dots + \mathbf{X}^{g-1} \quad (2.4.2)$$

ilmoittaa kaikkien erimittaisten kulkujen lukumäärät solmuparien välillä. Tästä summatrissiin saadaan saavutettavuusmatriisi $\mathbf{X}^{[R]}$, kun matriisin $\mathbf{X}^{[\Sigma]}$ nolasta poikkeavat alkiot merkitään ykkösiksi.

Saavutettavuusmatriisi on määritettävissä myös Warshallin algoritmilla (Ruohonen, 2006, 49) laskentatehokkaammin suurille verkostoille. Warshallin algoritmi pätee erityisesti suunnatuille graafeille, mutta toimii myös suuntaamattomille graafeille, kun solmuja yhdistävät kaaret tulkitaan vastakkaisuuntaisten nuolten pareina.

Warshallin algoritmilla graafin $\mathcal{G}(\mathcal{N}, \mathcal{L})$ saavutettavuusmatriisi $\mathbf{X}^{[R]}$ määritetään graafia \mathcal{G} kuvaavan sosiomatriisin \mathbf{X} avulla. Saavutettavuusmatriisi $\mathbf{X}^{[R]} = x_{ij}^{[R]}$, missä

$$x_{ij}^{[R]} = \begin{cases} 1, & \text{jos } \mathcal{G} \text{ :ssä on suunnattu } n_i\text{-}n_j\text{-polku} \\ 0, & \text{muuten.} \end{cases}$$

Warshallin algoritmi konstruoi $g \times g$ -matriisien jonon $\mathbf{E}_1, \dots, \mathbf{E}_g$, missä

1. \mathbf{E}_i :n alkiot ovat nollia ja ykkösiä
2. $\mathbf{E}_i \leq \mathbf{E}_{i+1}$ ($i = 0, \dots, n-1$) (vertailu alkioittain)
3. $\mathbf{E}_0 = \mathbf{X}$ (mikäli \mathbf{X} on arvotettu, saadaan \mathbf{E}_0 vaihtamalla \mathbf{X} :n positiiviset alkiot ykkösiksi)
4. $\mathbf{E}_g = \mathbf{X}^{[R]}$

Warshallin algoritmi pseudokoodina:

```

procedure Warshall
begin
   $\mathbf{E} := \mathbf{E}_0$ 
  for  $i := 1$  to  $n$  do
    for  $j := 1$  to  $n$  do
      if  $(\mathbf{E})_{ji} = 1$  then for  $k := 1$  to  $n$  do
         $(\mathbf{E})_{jk} := \max((\mathbf{E})_{jk}, (\mathbf{E})_{ik})$ 
      fi
    od
  od
end

```

Algoritmissa käytettyä maksimointioperaatiota kutsutaan usein *Boolean summaksi*, joka määritellään taulukossa 2.3 esitetyllä tavalla. Liitteessä A on esitetty Warshallin algoritmin toteutus Matlab-koodina.

Taulukko 2.3: Boolean summa.

max	0	1
0	0	1
1	1	1

2.4.3 Etäisyys ja geodeesi

Geodeesit eli solmujen lyhimmät etäisyydet esitetään usein etäisyysmatriisiin (*a distance matrix*) avulla. Etäisyysmatriisin alkiot $d(i, j)$ ilmoittavat solmujen n_i ja n_j välisen lyhimmän etäisyyden.

Lause 2.4.3. *(Di)graafin geodeesit löytyvät tarkastelemalla sosiomatriisia \mathbf{X} ja sen potenssimatriiseja $\mathbf{X}^2, \mathbf{X}^3, \dots, \mathbf{X}^{g-1}$ se.,*

$$d(i, j) = \min_{x_{ij}^{[p]} > 0} p.$$

Todistus. Koska sosiomatriisin \mathbf{X} potenssimatriisin \mathbf{X}^p alkiot x_{ij}^p ilmoittavat solmujen n_i ja n_j välisten p -mittaisten kulkujen lukumäärän, on solmujen n_i ja n_j välisen lyhyimmän etäisyyden pituus yhtä suuri kuin sen potenssimatriisin eksponentti, jolla alkio $x_{ij}^p \neq 0$, eli ts.

$$d(i, j) = \min_{x_{ij}^{[p]} > 0} p$$

□

Verkoston halkaisija on yhtä suuri kuin suurin verkostosta löytyvä geodeettinen etäisyys, eli ts. halkaisijan arvo on yhtä suuri kuin etäisyysmatriisin alkiodien maksimi, $\max[d(i, j)]$.

Geodeesi: Dijkstran algoritmi

Seuraavaksi esitellään lyhyesti geodeesin määrittämiseen ns. Dijkstran algoritmi. Esitys pohjautuu teoksessa Ruohonen (2006, 58-59) esitettyyn teoriaan.

Tarkastellaan yksinkertaista digraafia. Suuntaamaton graafi voidaan korvata digraafilla, kun sen kaarien tilalle otetaan vastakkaissuuntaisten nuolten parit. Merkitään nuolen $l_k = \langle n_i, n_j \rangle$ painoa (joka dikotomisilla (di)graafeilla on 0 tai 1 ja arvotetulla graafeilla jokin reaaliluku, ks. luku 2.5) $\alpha(l_k)$:lla. Dijkstran algoritmi ei päde negatiivisille painoille.

Dijkstran algoritmi käyttää pisteiden merkintää, joka voi olla pysyvä tai väliaikainen. Merkitään pisteen n_i merkkiä $\beta(n_i)$:lla ja määritellään

$$\gamma(n_i) = \begin{cases} 1, & \text{jos merkintä } \beta(n_i) \text{ on pysyvä} \\ 0, & \text{jos merkintä } \beta(n_i) \text{ on väliaikainen.} \end{cases}$$

Pysyvä merkki $\beta(n_i)$ ilmaisee kevyimmän suunnatun n_i - n_k -polun painon. Väliaikainen merkki $\beta(n_i)$ puolestaan antaa tälle painolle ylärajan (voi olla ∞).

Tarvittaessa otetaan käyttöön myös merkintä

$$\pi(n_i) = \begin{cases} \text{Pisteen } n_i \text{ edeltäjä (kevyimmällä) suunnatulla } n_i - n_k \text{-polulla,} \\ \text{jos sellainen on} \\ 0, & \text{muuten,} \end{cases}$$

jota käyttäen voidaan konstruoida ko. kevyin suunnattu polku, eli geodeesi.

Dijkstran algoritmi

1. Asetetaan $\beta(n_k) \leftarrow 0$ ja $\gamma(n_k) \leftarrow 1$.
Muille pisteille n_i asetetaan $\beta(n_i) \leftarrow \infty$ ja $\gamma(n_i) \leftarrow 0$.
Kaikille pisteille n_i asetetaan $\pi(n_i) \leftarrow 0$.
Asetetaan vielä $n_l \leftarrow n_k$.
2. Jokaiselle nuolelle $l_m = \langle n_l, n_i \rangle$, missä $\gamma(n_i) = 0$, asetetaan

$$\beta(n_i) \leftarrow \beta(n_l) + \alpha(l_m) \quad ja$$

$$\pi(n_i) \leftarrow n_l,$$

$$\text{jos } \beta(n_i) > \beta(n_l) + \alpha(l_m).$$

3. Etsitään piste n_i^* , jolle $\gamma(n_i^*) = 0$, $\beta(n_i^*) < \infty$ ja

$$\beta(n_i^*) = \min_{\gamma(n_i)=0} \{\beta(n_i)\}.$$

Asetetaan

$$\gamma(n_i^*) \leftarrow 1 \quad ja$$

$$n_l \leftarrow n_i^*.$$

Ellei mainitunlaista pistettä n_i^* ole, lopetetaan (suunnattuja n_k - n_m -polkujakaan ei silloin ole).

4. Jos $n_l \neq n_m$, palataan kohtaan 2.
5. Lopetetaan.

Algoritmin todistusta ei tässä kohtaa esitetä, mutta Ruohonen (2006, 59-60) esittää teoksessaan yhden tavan, jolla algoritmi voidaan nähdä oikeaksi.

2.4.4 Solmun aste

Suuntaamattomille verkostoille solmujen asteluvut ovat helposti laskettavissa sosiomatriisin \mathbf{X} tai insidenssimatriisin \mathbf{I} avulla.

Lause 2.4.4. *Suuntaamattoman verkoston solmun n_i asteluvulle pätee*

$$d(n_i) = \sum_{j=1}^g x_{ij} = \sum_{i=1}^g x_{ij}$$

sekä

$$d(n_i) = \sum_{j=1}^L I_{ij}.$$

Todistus. Sosiomatriisissa rivillä on merkitty 1:llä, jos saraketta vastaava solmu on liittynyt kaarella riviä vastaavaan solmuun. Nyt siis solmun n_i asteluku $d(n_i)$ saadaan sosiomatriisin rivisummana tai sarakesummana, koska matriisi on symmetrinen, eli ts.

$$d(n_i) = \sum_{j=1}^g x_{ij} = \sum_{i=1}^g x_{ij}.$$

Insidenssimatriisissa puolestaan rivillä on merkitty 1:llä, jos kaari on liittynyt solmuun ja 0:lla, jos kaari ei ole liittynyt solmuun. Nyt siis solmun n_i asteluku $d(n_i)$ saadaan insidenssimatriisin rivisummana, eli

$$d(n_i) = \sum_{j=1}^L I_{ij}.$$

□

Suunnatuille verkostoille solmujen venti- ja tuontiluvut (*outdegree*, *indegree*) ovat helposti laskettavissa sosiomatriisin \mathbf{X} avulla.

Lause 2.4.5. *Suunnatun verkoston solmun n_i vientiluvulle pätee*

$$d_O(n_i) = \sum_{j=1}^g x_{ij}.$$

Todistus. Suunnatun verkoston sosiomatriisin rivillä on 1, jos kyseistä riviä vastaavasta solmusta lähtee nuoli saraketta vastaavaan solmuun. Nyt siis solmun n_i vientiluku $d_O(n_i)$ saadaan sosiomatriisin rivisummana, eli

$$d_O(n_i) = \sum_{j=1}^g x_{ij}.$$

□

Lause 2.4.6. *Suunnatun verkoston solmun n_i tuontiluvulle pätee*

$$d_I(n_i) = \sum_{j=1}^g x_{ij}.$$

Todistus. Sosiomatriisin sarakkeessa on 1, jos saraketta vastaavaan solmuun tulee nuoli riviä vastaavasta solmusta. Nyt siis solmun n_i tuontiluku $d_I(n_i)$ saadaan sosiomatriisin sarakesummana, eli

$$d_I(n_i) = \sum_{j=1}^g x_{ij}.$$

□

2.4.5 Tiheys

Lause 2.4.7. *Suuntaamattomassa verkostossa, jossa on g toimijaa, tiheydelle Δ pätee*

$$\Delta = \frac{2 \sum_{i=1}^g \sum_{j=1}^g x_{ij}}{g(g-1)}.$$

Todistus. Verkoston tiheys määriteltiin verkostossa olemassa olevien solmujen välisten yhteyksien summan ja verkoston kaikkien mahdollisten solmujen välisten yhteyksien summan välisenä suhteena. Suuntaamattomassa verkostossa, jossa on g toimijaa, voi olla enintään $g(g-1)$ toimijaparien välistä suoraa yhteyttä, jolloin jokainen toimija on suorassa yhteydessä jokaiseen muuhun toimijaan. Toisaalta tällöin jokaista solmuparia yhdistää kaksi kaarta, jolloin rinnakkaiset kaaret voidaan poistaa. Täten verkoston kaarten maksimimäärä on

$$\frac{g(g-1)}{2}.$$

Verkoston sosiomatriisissa on merkitty 1:llä, mikäli toimijaparin välillä vallitsee yhteys ja 0:lla, jos toimijaparin väliltä puuttuu yhteys. Nyt siis olemassa olevien yhteyksien summa saadaan yksinkertaisesti sosiomatriisin kaikkien alkioiden summasta, eli tiheydelle Δ saadaan

$$\Delta = \frac{\sum_{i=1}^g \sum_{j=1}^g x_{ij}}{\frac{g(g-1)}{2}} = \frac{2 \sum_{i=1}^g \sum_{j=1}^g x_{ij}}{g(g-1)}.$$

□

Lause 2.4.7 pätee myös arvotetuille suuntaamattomille verkostoille (ks. luku 2.5). Kuitenkin on huomattava, ettei arvotetuilla verkostoilla tiheyden arvo aina asetu välille $[0,1]$, kuten dikotomisilla verkostoilla.

Lause 2.4.8. *Suunnatussa verkostossa, jossa on g toimijaa, tiheydelle Δ pätee*

$$\Delta = \frac{\sum_{i=1}^g \sum_{j=1}^g x_{ij}}{g(g-1)}.$$

Todistus. Vastaavasti kuin edellä tiheys vastaa verkostossa olemassa olevien solmujen välisten yhteyksien summan ja verkoston kaikkien mahdollisten solmujen välisten yhteyksien summan välistä suhdetta. Suunnatussa g toimijan verkostossa voi olla enintään $g(g-1)$ toimijaparien välistä suoraa yhteyttä, mikä vastaa verkoston yhteyksien maksimimäärää.

Edelleen verkoston olemassa olevien yhteyksien summa saadaan yksinkertaisesti sosisomatriisin kaikkien alkoiden summana, joten tiheydelle Δ saadaan

$$\Delta = \frac{\sum_{i=1}^g \sum_{j=1}^g x_{ij}}{g(g-1)}.$$

□

2.5 Arvotettujen verkostojen matriisit

Edellä esitetyt verkostot ovat olleet dikotomisia, eli niissä on otettu huomioon ainoastaan yhteyksien olemassaolo (tai olemassaolemattomuus). Arvotetuilla verkostoilla solmujen n_i ja n_j väliselle yhteydelle l_k voidaan liittää arvo tai paino v_k . Tämän avulla yhteydet voidaan arvottaa keskenään, sillä usein sosiaalisissa verkostoissa eri toimijaparien väliset yhteydet eivät ole keskenään samanlaisia tai saman tasoisia.

Määritelmä 2.5.1. Arvotetun verkoston sosisomatriisin \mathbf{X} alkio

$$x_{ij} = v_k,$$

missä v_k on yhteyden $l_k = \langle n_i, n_j \rangle$ paino.

Arvotetuille verkostoille pätevät käytännössä samat ominaisuudet ja määritelmät kuin edellä on (di)graafeille esitetty.

2.6 SNA-tunnusluvut

SNA-tunnuslukujen avulla saadaan selville sellaista tietoa verkoston rakenteesta ja toimijoista, jota ei näe suoraan verkoston sosisomatriisista, graafista tai muista vi-

sualisoinneista. Edellä esitetyt määritelmät ovat suurelta osin välttämättömiä seuraavien tunnuslukujen määrittämiseksi.

Seuraavassa esitettyjä teorioita ovat käsitelleet Wasserman ja Faust (1994), mutta heidän lisäksi myös mm. Knoke ja Yang (2008), Scott (2000), Hanneman ja Riddle (2005) ja Johanson, Mattila, ja Uusikylä (1998).

2.6.1 Suuntaamattomien verkostojen tunnusluvut: Keskeisyys

Suuntaamattomille verkostoille ja sen toimijoille voidaan määrittää erilaisia keskeyksyyksiä (*centrality*):

- Keskeyksyyssaste (*degree centrality*)
- Läheisyys (*closeness centrality*)
- Välillisyyys (*betweenness centrality*)

Lisäksi voidaan määrittää toimijan informaatiokeskeyksyyys (*information centrality*), mutta tämän tunnusluvun hyödyllisyys web-pohjaisten sosiaalisten verkostojen analysoinnissa ei ole niin merkittävä kuin edellä lueteltujen kolmen keskeyksyyden tunnusluvun.

Suuntaamattomissa verkostoissa keskeyksinen toimija on osallisena monissa yhteyksissä. Keskeyksyyden kannalta ei ole väliä, onko toimija lähettänyt vai vastaanottanut yhteyden.

Keskeyksyyys on verkoston toimijaa n_i kuvaava tunnusluku, kun taas koko verkostolle yhteisesti voidaan määrittää keskittyneisyys (*centralization*), joka on verkostosta toiseen vertailtava tunnusluku. Keskittyneisyyden avulla voidaan kuvata, missä määrin yksittäiset toimijat hallitsevat muiden välistä kanssakäymistä yleisesti koko verkoston tasolla.

Keskeyksyydet eivät riipu verkoston yhteyksien suunnasta, vaan ainoastaan tarkasteltavasta toimijasta ja siihen liittyneistä yhteyksistä ja/tai koko verkostosta yleensä. Suunnatuissa verkostoissa keskeyksyyden tilalla käytetään käsitettä arvostus (*prestige*), joka huomioi yhteyksien suunnan. Samalla erotetaan toisistaan käsitteet *olla arvostettu* ja *arvostaa*.

Keskeyksyyssaste

Toimijan n_i keskeyksyyssaste $C_D(n_i)$ kertoo, kuinka monta suoraa yhteyttä toimijalla on muihin toimijoihin (= asteluku $d(n_i)$). Toimijan aste ei itsessään ole kuitenkaan yleisesti vertailtava tunnusluku, mutta kun asteluku skaalataan verkoston tasolla lukualueelle $[0,1]$, voidaan toimijoiden keskeyksyyssasteita vertailla eri verkostojen välillä.

Määritelmä 2.6.1. Suuntaamattoman verkoston, jossa on g toimijaa, solmun n_i normeerattu keskeisyysaste $C'_D(n_i)$ määritellään

$$C'_D(n_i) = \frac{d(n_i)}{g-1}.$$

Verkoston normeerattujen keskeisyysasteiden varianssi määritellään

$$S_D^2 = \frac{\sum_{i=1}^g (C'_D(n_i) - \overline{C_D})^2}{g}, \quad (2.6.2)$$

missä $\overline{C_D}$ on normeerattujen keskeisyysasteiden keskiarvo ($\overline{C_D} = \sum_{i=1}^g (C'_D(n_i)/g)$).

Keskeisyysaste määritellään myös suunnatuille verkostoille, jolloin käsitellään erikseen vientikeskeisyyttä (*outdegree centrality*) ja tuontikeskeisyyttä (*indegree centrality*). Tällöin tulee siis tarkastella solmun n_i vienti- ja tuontilukuja erikseen ja määrittää niistä edellä esitetyllä tavalla määritelmää 2.6.1 soveltaen normeeratut vienti- ja tuontikeskeisyydet.

Verkoston keskittyneisyysaste (*group degree centralization*) kuvaa verkoston toimijoiden keskeisyysasteen vaihtelevuutta.

Määritelmä 2.6.3. Verkoston keskittyneisyysaste määritellään

$$C_D = \frac{\sum_{i=1}^g [C_D(n^*) - C_D(n_i)]}{\max \sum_{i=1}^g [C_D(n^*) - C_D(n_i)]} = \frac{\sum_{i=1}^g [C_D(n^*) - C_D(n_i)]}{(g-1)(g-2)},$$

missä $C_D(n^*) = \max\{C_D(n_i)\}$.

Läheisyys

Läheisyys $C_C(n_i)$ on toimijan n_i lyhyimpien polkujen (geodeesien) summa c_i kaikkiin verkoston muihin toimijoihin n_j . Se kertoo siis kuinka lähellä toimija on verkoston muihin toimijoihin nähden. Itsessään summa ei ole vertailtava tunnusluku eri verkostojen välillä, vaan jokaista solmua vastaava läheisyyden arvo tulee normeerata alueelle $[0,1]$.

Määritelmä 2.6.4. Suuntaamattoman verkoston, jossa on g toimijaa, solmun n_i normeerattu läheisyys määritellään

$$C'_C(n_i) = \frac{g-1}{c_i},$$

missä $c_i = \sum_{j=1}^g d(i, j)$ ja $d(i, j)$ on solmujen n_i ja n_j geodeesi.

Verkoston normeerattujen läheisyyksien varianssi määritellään

$$S_C^2 = \frac{\sum_{i=1}^g (C'_C(n_i) - \overline{C_C})^2}{g}, \quad (2.6.5)$$

missä $\overline{C_C}$ on normeerattujen läheisyyksien keskiarvo ($\overline{C_C} = \sum_{i=1}^g (C'_C(n_i)/g)$).

Verkoston läheisyyden keskittyneisyys (*group closeness centralization*) kuvaa sitä, kuinka läheisiä verkoston toimijat ovat toisiinsa nähden koko verkoston tasolla.

Määritelmä 2.6.6. Verkoston läheisyyden keskittyneisyys määritellään

$$C_C = \frac{\sum_{i=1}^g [C'_C(n^*) - C'_C(n_i)]}{\max \sum_{i=1}^g [C'_C(n^*) - C'_C(n_i)]} = \frac{\sum_{i=1}^g [C'_C(n^*) - C'_C(n_i)]}{\frac{(g-1)(g-2)}{2g-3}},$$

missä $C'_C(n^*) = \max\{C'_C(n_i)\}$.

Välillisuus

Toimijan n_i välillisuus $C_B(n_i)$ puolestaan mittaa, kuinka monen toimijaparin välisen lyhyimmän polun varrelle toimija sijoittuu. Toimijan välillisuus on merkityksellinen tunnusluku mm. tutkittaessa verkoston toiminnan tehokkuutta.

Määritelmä 2.6.7. Suuntaamattoman verkoston, jossa on g toimijaa, solmun n_i normeerattu välillisuus $C'_B(n_i)$ määritellään

$$C'_B(n_i) = \frac{\sum_{j < k} g_{jk}(n_i)}{\frac{(g-1)(g-2)}{2}},$$

missä g_{jk} on toimijoita n_j ja n_k yhdistävien geodeesien lukumäärä.

Normeerattu välillisuus saa arvoja välillä $[0,1]$. Mitä suurempi välillisyyden arvo on, sitä useammalle eri toimijaparin n_j ja n_k väliselle geodeesille toimija n_i sijoittuu.

Verkoston normeerattujen välillisyyksien varianssi määritellään

$$S_B^2 = \frac{\sum_{i=1}^g (C'_B(n_i) - \overline{C_B})^2}{g}, \quad (2.6.8)$$

missä $\overline{C_B}$ on normeerattujen välillisyyksien keskiarvo ($\overline{C_B} = \sum_{i=1}^g (C'_B(n_i)/g)$).

Verkoston välillisyyden keskittyneisyys (*group betweenness centralization*) kuvaa verkoston toimijoiden välillisyyden heterogeenisuutta eli toimijoiden välillisyyden vaihtelevuutta.

Määritelmä 2.6.9. Verkoston välillisyyden keskittyneisyys määritellään

$$C_D = \frac{\sum_{i=1}^g [C'_B(n^*) - C'_B(n_i)]}{\max \sum_{i=1}^g [C'_B(n^*) - C'_B(n_i)]} = \frac{\sum_{i=1}^g [C'_B(n^*) - C'_B(n_i)]}{(g-1)},$$

missä $C'_B(n^*) = \max\{C'_B(n_i)\}$.

2.6.2 Suunnattujen verkostojen tunnusluvut: Arvostus

Arvostuksen tunnusluvuissa tarkastellaan nimenomaan yhteyksien vastaanottamista eli arvostuksen kohteena olemista. Verkoston toimijoille voidaan määrittää seuraavat arvostukset:

- Arvostusaste (*actor degree prestige*)
- Arvostusläheisyys (*actor proximity prestige*)
- Arvoasema (*actor status prestige, actor rank prestige*)

Koko verkoston tasolla voidaan määrittää verkostoa kuvaavia arvostuksen tunnuslukuja arvioimalla kunkin arvostuksen keskiarvoja ja variansseja. Näistä keskiarvostusläheisyys ja arvostusläheisyyden varianssi ovat mielekkäimpiä verkostanalyysin kannalta.

Arvostusaste

Toimijan n_i arvostusaste $P_D(n_i)$ määritellään yksinkertaisesti toimijaan kohdistuneiden yhteyksien summana (vrt. suuntaamattoman verkoston toimijan n_i keskeisyysaste, joka määritellään toimijaa kuvaavan solmun astelukuna $d(n_i)$).

Määritelmä 2.6.10. Suunnatun verkoston toimijan n_i arvostusaste määritellään

$$P_D(n_i) = d_I(n_i) = \sum_j x_{ij} = x_{+i},$$

missä $d_I(n_i)$ on solmun n_i tuontiluku ja x_{+i} on verkoston sosiomatriisin \mathbf{X} sarakesumma sarakkeesta i .

Jotta arvostusasteet eri verkostojen välillä olisivat verrattavissa keskenään, määritellään normeerattu arvostusaste $P'_D(n_i)$.

Määritelmä 2.6.11. Suunnatun verkoston toimijan n_i normeerattu arvostusaste määritellään

$$P'_D(n_i) = \frac{x_{+i}}{g - 1},$$

missä g on verkoston toimijoiden lukumäärä.

Normeerattu arvostusaste saa arvoja väliltä $[0,1]$. Sitä arvostetumpi toimija on, mitä suurempi on sen arvostusaste.

Arvostusläheisyys

Toimijan n_i vaikutusjoukko (*influence domain*) koostuu niistä toimijoista n_j , joista toimija n_i on saavutettavissa. Toimijalle n_i voidaan määrittää luku I_i , joka ilmoittaa, kuinka moni toimija n_j voi saavuttaa toimijan n_i . Tämä luku I_i toimijalle n_i saadaan yksinkertaisesti saavutettavuusmatriisin $\mathbf{X}^{[R]}$ sarakesummana sarakkeesta i .

Toimijan n_i vaikutusjoukon toimijoiden lukumäärän I_i avulla voidaan määrittää vaikutusjoukon toimijoiden keskimääräinen etäisyys toimijasta n_i

$$d(n_j, n_i)_{ave} = \frac{\sum_j d(n_j, n_i)}{I_i}. \quad (2.6.12)$$

Toimijan n_i arvostusläheisyys $P_P(n_i)$ määritellään vaikutusjoukon osuuden koko verkoston toimijajoukosta suhteena vaikutusjoukon toimijoiden keskimääräiseen etäisyyteen toimijasta n_i .

Määritelmä 2.6.13. Suunnatun verkoston toimijan n_i arvostusläheisyys määritellään

$$P_P(n_i) = \frac{I_i/(g-1)}{\sum_j d(n_j, n_i)/I_i}.$$

Arvostusläheisyys saa arvoja välillä $[0,1]$. Jos toimijalla n_i on suora yhteys jokaiseen verkoston muuhun toimijaan, saa arvostusläheisyys arvon yksi. Jos taas toimijaan n_i ei kohdistu yhtään suoraa yhteyttä, vaikutusjoukon toimijoiden lukumäärä I_i on nolla ja arvostusläheisyys määritellään nolaksi.

Yleisesti erilaisia arvostusindeksejä on vaikea yleistää kuvaamaan koko verkostoa, eli sitä millä tasolla toimijoiden välinen arvostus verkoston sisällä on. Arvostusläheisyydestä voidaan kuitenkin yksinkertaisesti määrittää verkostotason tunnuslukuja arvostukseen liittyen. Keskiarvostusläheisyys kuvaa yleisesti sitä, kuinka läheisesti toimijat arvostavat toinen toistaan verkoston sisällä.

Määritelmä 2.6.14. Suunnatussa verkostossa, jossa on g toimijaa, keskiarvostusläheisyys määritellään

$$\overline{P_P} = \sum_{i=1}^g \frac{P_P(n_i)}{g}.$$

Edelleen verkoston arvostusläheisyyksien varianssi on koko verkostoa kuvaava tunnusluku

$$S_P^2 = \sum_{i=1}^g \frac{(P_P(n_i) - \overline{P_P})^2}{g}. \quad (2.6.15)$$

Arvoasema

Edellä esitetyt arvostuksen tunnusluvut huomioivat vain toimijoita ja niihin kohdistuvia yhteyksiä. Toimijan n_i arvoasema on riippuvainen häneen päin yhteydessä olevien toimijoiden n_j arvoasemista. Edelleen taas toimijoiden n_j arvoasemat ovat riippuvaisia näihin päin yhteydessä olevien toimijoiden n_k arvoasemista jne. Arvoaseman määrittelyn taustalla on ajatus siitä, että toimijan n_i arvoasema on häneen päin suoraan yhteydessä olevien toimijoiden arvoasemien funktio.

Sosiomatriisin \mathbf{X} sarakkeessa on merkitty luvulla yksi, jos riviä kuvaava toimija on yhteydessä saraketta kuvaavaan toimijaan, ja nollalla, jos yhteyttä ei ole. Täten verkostossa, jossa on g toimijaa, toimijan n_i arvoasema voidaan esittää sosiomatriisin \mathbf{X} toimijaa kuvaavan sarakkeen alkioden ja niitä vastaavien toimijoiden arvoasemien lineaarikombinaationa, ts.

$$P_R(n_i) = x_{1i}P_R(n_1) + x_{2i}P_R(n_2) + \dots + x_{gi}P_R(n_g). \quad (2.6.16)$$

Matemaattisesti koko verkoston toimijoiden arvoasemat voidaan esittää matriisiyhtälönä

$$\mathbf{p} = \mathbf{X}^T \mathbf{p}, \quad (2.6.17)$$

missä $\mathbf{p} = (P_R(n_1), P_R(n_2), \dots, P_R(n_g))^T$ (pystyvektori) ja \mathbf{X} on sosiomatriisi (ja \mathbf{X}^T sen transpoosi). Tästä matriisiyhtälöstä tulee siis ratkaista arvoasemavektori \mathbf{p} .

Matriisiyhtälö voidaan esittää muodossa

$$(\mathbf{I} - \mathbf{X}^T)\mathbf{p} = \mathbf{0}, \quad (2.6.18)$$

missä \mathbf{I} on identiteettimatriisi ja $\mathbf{0}$ g -pituinen vakio pystyvektori, jonka alkiot ovat nollia. Tämä yhtälö on identtinen nk. karakteristisen yhtälön kanssa. Karakteristista yhtälöä käytetään matriisien ominaisarvojen etsimiseen.

Tämän karakteristisen yhtälön ratkaisuproseduuriin ei tässä perehdytä sen tarkemmin. Erään ratkaisumallin esittää Katz (1953). Sitä ei tässä esitetä, mutta sen lähtökohtana on, että sosiomatriisin \mathbf{X} sarakesummat normeerataan vastaamaan arvoa 1. Edelleen \mathbf{X}^T :n yhden ominaisarvon pakotetaan olevan 1 ja tätä ominaisarvoa vastaavaksi ominaisvektoriksi pakotetaan \mathbf{p} .

2.7 Sosiaaliset verkostotyyppit

Edellä on esitelty SNA-menetelmiä sosiaalisille verkostoille, joissa toimijat ovat keskenään samanlaisia (ihmisiä tms. toimijoita, joilla oletetaan olevan samanlaiset ominaisuudet). Näiden yksilöiden muodostamia verkostoja sanotaan yksimoodisiksi (*one mode*), verkostoiksi.

Kun käsitellään verkostoja, joissa on toimijoita, esim. ihmisiä, ja toiminnan kohteita, esim. tapahtumia (*events*), puhutaan ns. kaksimoodisesta (*two mode*) verkostosta. Kaksimoodisessa verkostossa toimijat voivat olla yhteydessä sekä keskenään että toiminnan kohteisiin.

2.7.1 Yksimoodinen verkosto

Yksimoodisten verkostojen ominaisuudet ja SNA-menetelmät ovat suoraan niitä, joita on edellä esitelty. Verkoston mallina toimii (di)graafi, jolle voidaan määrittää vastaava sosiomatriisi. Verkostanalyysin kaikki laskennalliset ominaisuudet ovat voimassa yksimoodisille verkostoille.

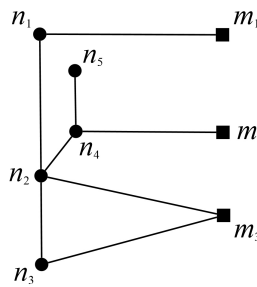
2.7.2 Kaksimoodinen verkosto

Tässä esitellään lyhyesti kaksimoodisten verkostojen mallit. Kaksi-moodista verkostoa voidaan mallintaa yhtäläillä graafina ja matriisina, aivan kuten yksimoodista verkostoa. Kuitenkin kaksimoodisuus luo rajoitteita SNA:lle.

Kaksimoodisen verkoston graafi

Kaksimoodisen verkoston graafissa on kaksi eri solmujoukkoa, joista toinen edustaa sosiaalisissa verkostoissa toimijoita, eli ihmisiä, ja toinen tapahtumia. Verkostossa merkitään toimijoiden joukkoa $\mathcal{N} = \{n_1, n_2, \dots, n_g\}$ ja tapahtumien joukkoa $\mathcal{M} = \{m_1, m_2, \dots, m_h\}$. Näitä yhdistää kaarien tai nuolien joukko $\mathcal{L} = \{l_1, l_2, \dots, l_L\}$.

Huomattavaa on nyt, että yleisesti toimijat voivat olla yhteydessä toistensa kanssa, mutta myös yhteydessä tapahtumiin. Tapahtumat eivät kuitenkaan ole vastavuoroisessa yhteydessä keskenään. Täten yhteyksiä voi siis olla toimijaparien välillä sekä toimijan ja tapahtuman välillä. Mikäli yhteyksiä ei ole toimijoiden välillä lainkaan, redusoituu tilanne vastaamaan malliltaan kaksijakoista graafia, jota on käsitelty aiemmin luvussa 2.2.10. Kuvassa 2.8 on esitetty esimerkki kaksimoodisen verkoston graafista.



Kuva 2.8: Kaksimoodisen verkoston graafi.

Kaksimoodisen verkoston matriisi

Kaksimoodiselle verkostolle voidaan määritellä sosiomatriisi \mathbf{X} se., solmujoukot \mathcal{N} ja \mathcal{M} yhdistetään yhdeksi joukoksi. Siinä toimijat luetellaan ennen tapahtumia ja tällöin sosiomatriisi \mathbf{X} on $(g + h) \times (g + h)$ neliömatriisi.

Tämä sosiomatriisi soveltuu hyvin verkoston mallintamiseen, mutta sen laskennalliset ominaisuudet kaksimoodisten verkoston analyysin kannalta jäävät vähäisiksi. Siksi määritelläänkin nk. jäsenyysmatriisi (*affiliation matrix*) \mathbf{A} , jossa jokainen sarakke vastaa tiettyä kaksimoodisen verkoston tapahtumaa m_i ja vastaavasti jokainen rivi vastaa tiettyä toimijaa n_i .

Määritelmä 2.7.1. Jäsenyysmatriisin \mathbf{A} alkiot a_{ij} määritellään

$$a_{ij} = \begin{cases} 1, & \text{jos toimija } n_i \text{ on yhteydessä tapahtumaan } m_j \\ 0, & \text{muuten.} \end{cases}$$

Jos, nyt tarkastelemme aiemmin esiteltyä kaksimoodisen verkoston sosiomatriisia \mathbf{X} , huomaamme, että se on muotoa

$$X = \begin{bmatrix} B & A \\ A^T & 0 \end{bmatrix},$$

missä \mathbf{B} on toimijajoukon \mathcal{N} $g \times g$ sosiomatriisi (yksimoodinen verkosto) ja $\mathbf{0}$ on $h \times h$ nollamatriisi.

Jäsenyysmatriisin \mathbf{A} avulla voidaan tutkia helposti tapahtumiin yhteydessä olevien toimijoiden määriä (\mathbf{A} :n sarakesummat) ja vastaavasti toimijan yhteyksien määriä tapahtumiin (\mathbf{A} :n rivisummat). Edelleen, voidaan tutkia mm. sitä ovatko tietyt toimijaparit yhteydessä samoihin tapahtumiin tai toimijoiden määriä suhteessa tiettyihin tapahtumapareihin. Näistä mm. Wasserman ja Faust (1994) antavat havainnollistavia esimerkkejä.

3. WEB-POHJAISET SOSIAALISET VERKOSTOPALVELUT JA NIIDEN SNA

Preece, Maloney-Krichmar ja Abras (2003) määrittelevät web-pohjaisen verkkoyhteisön (*a web community*) joukkona ihmisiä, jotka kokoontuvat yhteen verkon välityksellä jonkin yhteisen tarkoituksen myötä ja/tai tavoitteen saavuttamiseksi. Yhteisön yksilöitä ohjaavat yhteisössä vallitsevat normit ja säännöt (Preece et al., 2003). boyd ja Ellison (2007) puolestaan määrittelevät verkkoyhteisön sosiaalisen verkostoitumisen ja yksilöitä yhdistävien yhteyksien luomisen näkökulmasta. Tyypillisiä toimintoja tällaisissa sosiaalisissa verkkoyhteisöpalveluissa ovat yksilöidyt käyttäjäprofiilit, käyttäjien väliset yhteydet, sisällön kommentit ja yksityisviestimet. Lisäksi verkkoyhteisöissä voi olla useita lisäominaisuuksia (esim. sisällön jakamisen työkaluja, blogeja, viestimiä), jotka erottavat eri verkkoyhteisöt toisistaan.

Toisaalta sosiaalisia verkkoyhteisöpalveluita voidaan tarkastella teknisestä näkökulmasta. O'Reilly (2005) määrittelee käsitteen *Web 2.0*, joka viittaa aktiivisempiin ja loppukäyttäjääorientoituneempiin web-sovelluksiin, joissa keskitytään sosiaalisen sisällön hallintaan, tuottamiseen ja jakamiseen. Tyypillisiä teknisiä Web 2.0:n ratkaisuja ovat AJAX (*Asynchronous JavaScript and XML*), joka mahdollistaa korkeantason vuorovaikutteisten sovellusten toteutuksen, sekä web-syötteet, kuten RSS (*Really Simple Syndication*), joka tarjoaa ratkaisun mm. uutisvirtojen viemiselle ja tuomiselle eri verkkosovellusten välillä.

3.1 Sosiaalisen verkostopalvelun SNA

Sosiaalisten verkostojen tutkijat voivat hyödyntää rikasta verkostodataa tutkimalla nk. CMC-järjestelmiä (*Computer-mediated Communication Systems*), kuten sähköposteja, keskustelupalstoja, ääniviestejä, ja muita ryhmätoimintoja. Koska monitoroinnissa yleisesti tallennetaan tieto kaikista interaktioista CMC-järjestelmään, verkostodata sisältää tiedon kaikista vuorovaikutuksista verkoston toimijoiden välillä. Kuitenkin, tarkka toimintojen seuranta voi olla työlästä laskennallisen ajan ja tehokkuuden näkökulmasta katsottuna. Edelleen yhteyksien muuttuessa sisällön, ajan ja tarkoituksen suhteen, yksityiskohtainen data ei ole välttämättä relevanttia. (Knoke & Yang, 2008, 30.)

CMC-järjestelmiä tutkittaessa voidaan käsiteltävää dataa ja sen määrää vähentää tutkimuksen luonteen ja tulkinnan myötä. CMC-data tarkasti reflektoi mm. CMC-

järjestelmän toimintaa ja sisältöäkin, mikä mahdollistaa esim. toiminnan vastavuoroisuutta koskevien hypoteesien eksplisiittisen erottelun ja testaamisen (Knoke & Yang, 2008, 31). Vaikkakin CMC-data on tarkempaa kuin esim. kyselyin kerätty data, se ei välttämättä ole sen validimpaa. Se usein esittää ihmisten kommunikaation eri aspekteja (Rice, Borgman, Bednarski, & Hart, 1989).

Monien web-pohjaisten verkkoympäristöjen tekniikka mahdollistaa monipuolisen datan tallentamisen verkkoyhteisössä tapahtuvasta käyttäjätoiminnasta. Tietokantoihin tehtävillä kyselyillä voidaan yksinkertaisesti kerätä tietoa sekä yksittäisistä että useista toiminnoista joko koko verkoston toimijatasolla tai tarkastellen vain yhtä käyttäjää. Rakenteiseen XML (*Extensible Markup Language*) -dokumenttiin voidaan kerätä kaikki tieto verkoston toiminnasta yksilöitynä niin ajan, toiminnon tekijän kuin toiminnon laadun mukaan. Kokonaisuutena kaikista verkkopalvelun toiminnoista voidaan kerätä lokidataa, joka on XML-formaatissa helposti käsiteltävissä ja analysoitavissa.

Tätä dataa voidaan hyödyntää SNA:ssa, joka mahdollistaa rikkaan verkkoinformaation tuottamisen. Lokidataan perustuvan verkostanalyysin kautta voidaan tuottaa erilaisia verkostoa kuvaavia ja mallintavia visualisointeja. Visualisointeja voidaan hyödyntää sosiaalisen verkostanalyysin tukena mm. kun halutaan havaita muutoksia ja heikkoja signaaleja, luoda yleisnäkyymiä koko verkostoon ja tuoda verkoston kehitys näkyväksi (Huhtamäki, Nykänen, & Salonen, 2009a).

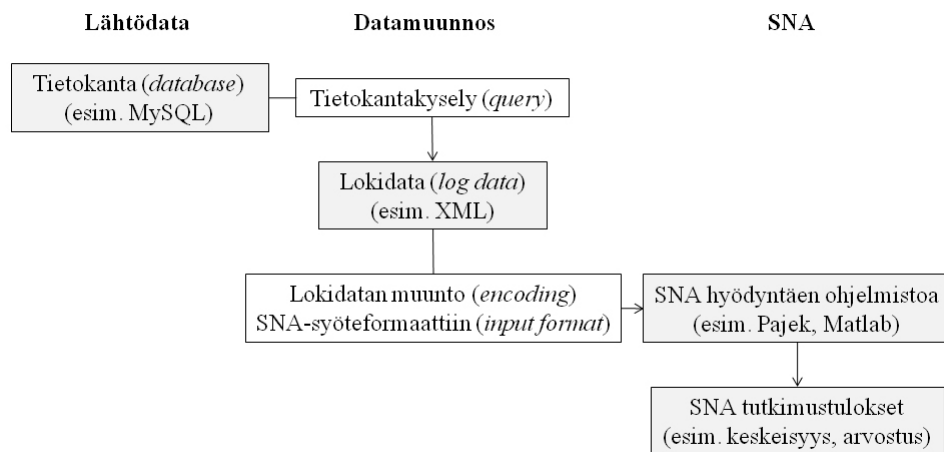
3.2 Sosiaalisen verkostopalvelun analyysiprosessi

Web-pohjaiset järjestelmät mahdollistavat laaja-alaisen ja yksityiskohtaisen lokidatan keruun ja tallentamisen. Kuten edellä mainittiin, on XML-muotoinen lokidata helposti käsiteltävissä vastaamaan useita verkostanalyysin tarpeita. Keskeistä on kuitenkin huomioida, että toteutettavaa analyysiä ohjaavat hypoteesit, käytävissä oleva lokidata, lokidatan rekonstruointi verkostodataksi sekä tiedon tulkinta osana analyysia. Koska web-pohjaiset sosiaaliset verkostopalvelut käsittävät laajoja verkostoja ja siten verkostodata käsittää verrattain paljon tietoa, on järkevää määrittellä verkostanalyysille johdonmukainen looginen prosessi, jota noudattamalla verkoston jokainen eri analyysi on paremmin vertailtavissa ja suhteutettavissa toisten analyysien kanssa.

Tällaisia loogisia prosesseja ei ole kirjallisuudessa juurikaan esitelty, sillä monesti prosessi riippuu tutkittavasta verkostosta itsessään, siitä saatavasta verkstodatasta sekä toteutettavasta tutkimuksesta. Seuraavassa kuitenkin esitellään yleinen malli, jolla voidaan toteuttaa useimpien web-pohjaisten sosiaalisten verkostojen analyysia, olettaen että data ja työvälineet ovat suoraan käytettävissä.

3.2.1 Verkostanalyysin looginen prosessi

Yleisesti verkostanalyysin loogisessa prosessissa on kolme vaihetta. Alussa on lähtödata, joka on yleisimmin tallennettu tietokantoihin (*database*, esim. MySQL - *My Structured Query Language*). Tämä tietokannoissa oleva data saadaan ulos tietokantakyselyillä (*query*) esim. XML-muodossa, ja se muunnetaan edelleen datamuunnosvaiheessa SNA-syötedataksi oikeassa syöteformaattissa (*input format*). Tämän jälkeen voidaan toteuttaa varsinainen verkostanalyysi siihen soveltuvien välineiden avulla. Tämä looginen prosessi on havainnollistettu kuvassa 3.1.



Kuva 3.1: Web-pohjaisen sosiaalisen verkoston analyysin looginen prosessi.

Jotta SNA:sta voitaisiin saada mahdollisimman automatisoitua, kannattaa tietokantakyselyssä sekä lokidatan (*log data*) muunnossa (*encoding*) hyödyntää automatisoituja ratkaisuja tämän tyyppisten toimintojen suorittamiseen. Erityyppisistä tietokannoista tehtäviin kyselyihin on mahdollista toteuttaa räätälöityjä työkaluja, joilla tietokantaan tallennettu lokidata voidaan tuottaa jatkokäsiteltäväksi XML-muodossa. Edelleen tämä XML-data voidaan muuntaa SNA-syötedataksi hyödyntäen muita työkaluja, joista tässä esimerkkinä mainittakoon Wille-ympäristö (ks. Nykänen, Salonen, Haapaniemi, ja Huhtamäki (2008)). Lokitiedon prosessointia on esitelty tarkemmin raportissa Huhtamäki, Nykänen, ja Salonen (2009b).

3.2.2 SNA-ohjelmistot

Verkostanalyysiin on olemassa useita eri käyttötarkoituksiin soveltuvia työkaluja. Lähinnä kaikki matematiikka- ja tilasto-ohjelmistot (esim. Matlab, Maple, SPSS) soveltuvat analyysin toteuttamiseen, kun syötedatana hyödynnetään matriiseja. Eri-tyyppisistä SNA-ohjelmistoista monipuolisin on Pajek.

Pajek (sloveniaa, hämähäkki) on Windows-ohjelma, joka mahdollistaa suurten

verkostojen analyysin ja visualisoinnin (ks. esim. Nooy, Mrvar, ja Batagelj (2005); Batagelj ja Mrvar (1998); Tampubolon (2004)). Ohjelma tarjoaa hyvät työkalut verkostanalyysiin, sisältäen yleisimmät SNA-menetelmillä tuotettavat tiedot (mm. keskeisyys). Edelleen, ohjelman avulla on mahdollista visualisoida suuria verkostoja sekä tasossa että 3-ulotteisessa avaruudessa.

Muista SNA-ohjelmistoista mainittakoon UCINET, joka on kaupallinen verkostanalyysiohjelma. Microsoft Excel 2007:lle on saatavilla NodeXL-laajennus, joka mahdollistaa sosiaalisten verkostojen visualisoinnin ja SNA:n yleensä, hyödyntäen Excelin ominaisuuksia. Verkostojen visualisointiin on saatavilla useita erityyppisiä ohjelmistoja ja ratkaisuja. Näistä mainittakoon NetDraw, Vizster, Commentrix ja SoNIA.

4. CASE: TTY-PIIRI - SOSIAALISEN VERKOSTOPALVELUN VERKOSTOANALYYSI

Tässä luvussa käsitellään case-tutkimusta, jossa edellä esitettyjä sosiaalisen verkostanalyysin teorioita ja menetelmiä sovellettiin Tampereen teknillisen yliopiston sisäisen verkostoitumispalvelun TTY-Piirin verkostanalyysissä. Tutkimuksen ensisijaisena tarkoituksena oli testata palveluun kehitetyn lokidatan keräystyökalulla tuotetun verkostodatan analyysiä käytännössä. Tavoitteena oli löytää ne teorit ja menetelmät, joilla kyseisen verkostopalvelun käyttäjäverkoston toimintaa ja muutoksia pystyttäisiin indikoimaan ja havainnollistamaan. Kuitenkaan tutkimuksessa ei pyritty täysimittaiseen verkostanalyysiin, sillä jo ennakkoon oli odotettavissa, että määritelty verkostodata ja sen analyysimenetelmät olivat vasta kehitysasteella.

Verkostanalyysin kanssa toteutettiin palvelun käytönseurantaa, ja tarkasteltiin onko palvelun käyttöaktiivisuudella suoraa vaikutusta palvelun sisällä tapahtuvaan käyttäjien väliseen verkostoitumiseen. Lisäksi tutkimus sidottiin osaksi palvelun kehittämistutkimustyötä, jossa mm. laadullisin tutkimusmenetelmin tutkittiin ja karotettiin korkeakoulukontekstiin sidotun verkostoitumispalvelun käyttöä motivoivia tekijöitä (Silius, Miilumäki, Huhtamäki, Tebest, Meriläinen & Pohjolainen, 2010a, 2010b).

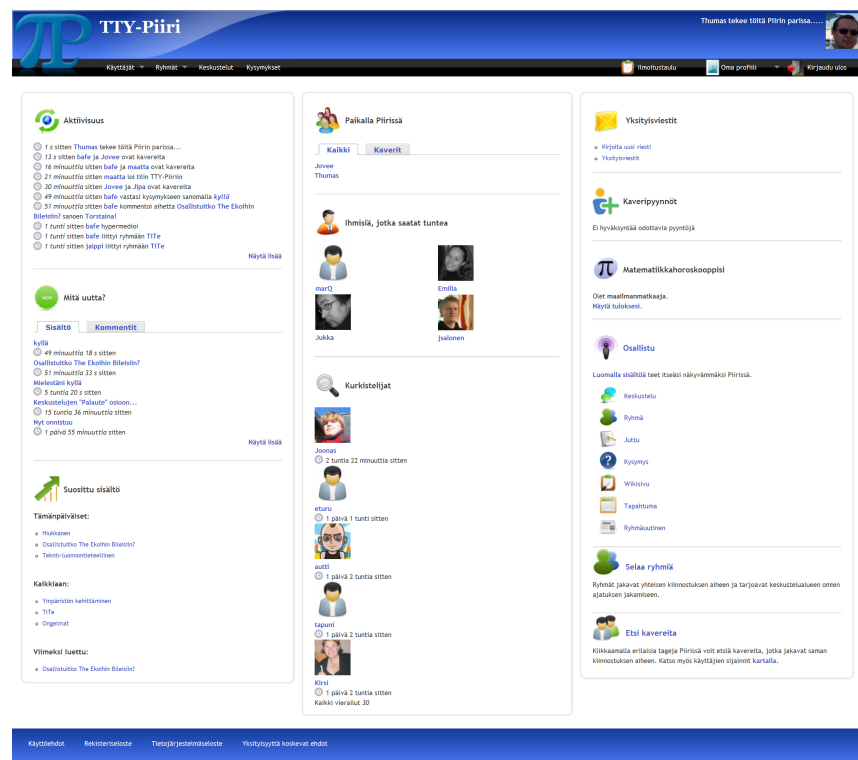
Seuraavaksi esitellään pääpiirteittäin tutkimus-casessa käsiteltävää sosiaalista verkostoitumispalvelua TTY-Piiriä. Tarkoitus ei ole kuvata palvelun teknisiä ratkaisuja liian yksityiskohtaisesti, vaan kuvauksessa keskitytään palvelun toiminnallisten ratkaisujen sekä sisältöjen kuvaukseen. Kuvaus pohjautuu pääpiirteittäin artikkeleissa Silius et al. (2010a, 2010b) julkaistuihin TTY-Piirin kuvauksiin.

4.1 TTY-Piiri

Sosiaalinen verkostopalvelu TTY-Piiri on Tampereen teknillisen yliopiston (TTY) opiskelijoiden web-pohjainen ympäristö, joka tarjoaa räätälöityjä ratkaisuja tehostamaan sosiaalista vuorovaikutusta TTY:lla. Palvelun ensisijainen tavoite on tarjota tehokas ja tiivis järjestelmä, joka auttaa TTY:n uusia opiskelijoita luomaan sosiaalisia suhteita keskenään ja TTY:n muiden opiskelijoiden, kuten tutoreiden, kanssa. Tämän lisäksi palvelun avulla on mahdollista luoda opiskeluryhmiä ja opintopiirejä opiskelun eri tarpeisiin. Tällaiset opintoja tukevat ratkaisut ovat osa palvelun kokonaisuutta.

TTY-Piiri pohjautuu avoimen lähdekoodin *Drupal*-sisällönhallintajärjestelmään (*CMS - Content Management System*) tai (*CMF - Content Management Framework*), joka tarjoaa perustyökalut erilaisten sisältöjen hallintaan, tuottamiseen ja editoimiseen. Erityisillä moduleilla järjestelmää voidaan räätälöidä vastaamaan erilaisia tarpeita.

TTY-Piirissä jokaisella palvelun käyttäjällä on oma personoitu ilmoitustaulu, joka koostaa kaikkea palvelun sisäistä toimintaa ja sisältöjä yhdelle sivulle systemaattisesti järjestetyllä tavalla reaaliajassa (ks. kuva 4.1). Kooste pitää sisällään tietoja käyttäjän omasta toiminnasta, lähipiirin, ts. palvelussa kavereiksi merkittyjen käyttäjien, toiminnasta sekä lähipiirin ulkopuolisten, ts. kaikkien muiden palvelun käyttäjien, toiminnasta. Lähipiirin ulkopuolisten toimijoiden aktiivisuuden esille tuominen mahdollistaa käyttäjää laajentamaan sosiaalista verkostoaan palvelun sisällä.



Kuva 4.1: TTY-Piirin ilmoitustaulu.

Sosiaalisena verkostopalveluna TTY-Piiri tarjoaa kaikki sosiaalisen median yleisimmät työkalut yksilöiden väliseen interaktioon. Jokaisella käyttäjällä on oma profiili, jossa voi kattavasti jakaa tietoa itsestään palvelun muille käyttäjille. Tyypillisten henkilötietojen (esim. nimi, ikä, sukupuoli, syntymäaika, jne.) lisäksi voi käyttäjä jakaa tietoa mm. harrastuksistaan, osaamisestaan ja kiinnostuksen kohteistaan, mikä mahdollistaa samankaltaisten käyttäjien etsimisen ja löytämisen, sekä täten verkostoitumisen myös entuudestaan täysin tuntemattomien ihmisten kanssa.

Jokainen käyttäjä voi luoda itselleen verkoston lisäämällä palvelun muita käyttäjiä kavereikseen. Kaverisuhteet ovat sosiaalisen verkostanalyysin kannalta mielekäs tutkimuskohde. TTY-Piirissä käyttäjien välinen kaverisuhde muodostuu, kun käyttäjä hyväksyy jonkin toisen käyttäjän hänelle osoittaman kaveripyynnön. Kaverisuhde on tällöin suoraan dikotominen yhteys toimijaparin välillä. Useat kaverisuhteet palvelun sisällä muodostavat käyttäjän lähipiirin, jonka toiminnasta koostetaan korkeammalla prioriteetilla tietoa käyttäjän ilmoitustaululle. Edelleen yksinkertaisella SNA-tarkastelulla voidaan tutkia palvelun verkostoa ja löytää nk. kaverin kavereita, joita voidaan edelleen suositella käyttäjälle liitettäväksi osaksi lähipiiriä.

TTY-Piiri sisältää ryhmiä, joita käyttäjät voivat luoda vapaasti. Ryhmä tarjoaa monipuolisia välineitä kollaboratiiviseen työskentelyyn ja yhteisen intressin jakavien käyttäjien väliseen interaktioon. Palvelun ryhmät ja ryhmiin liittyneet käyttäjät muodostavat kaksimoodisen verkoston, jota voidaan tarkastella ja visualisoida SNA-menetelmin.

Ryhmät tarjoavat yhden mahdollisuuden käyttäjien väliseen keskusteluun, mutta palvelun sisällä laajempaan ja yleisempään keskusteluun on tarkoitettu oma keskustelualue, joka toimii samalla tavalla, kuin yleisimmät internetin keskustelupalstat. Edelleen TTY-Piirissä on toiminnallisuus, joka mahdollistaa anonyymien kysymysten esittämisen ja niihin vastaamisen. Tämä on erityisen hyödyllinen opiskeluun liittyvissä asioissa, kun opiskelija haluaa esittää ns. tyhjän kysymyksen, joka kuitenkin saattaa olla ajankohtainen monelle muulle käyttäjälle. Asiaan vastauksen tietävä henkilö voi vapaasti vastata kysymykseen palvelun sisällä.

TTY-Piiri on jatkuvan tutkimus- ja kehitystyön kohteena. Palveluun toteutetaan uusia toiminnallisuuksia, jotka julkaistaan palveluun tarkasti suunniteltuina interventioina. Case-tutkimus pohjautuu syksyllä 2009 toteutettujen interventioiden vaikutusten tutkimiseen TTY-Piirin sosiaalisessa verkostossa. Tutkimuksessa keskeisessä roolissa ovat SNA-menetelmät ja sen tuloksia peilataan palvelusta peruskäytönseurannan kautta saatavaan tietoon.

4.2 Tutkimuskysymykset

TTY-Piirin käyttömotivaatiota koskevissa tutkimuksissa korostui selvästi palvelun sisäisen toiminnan ja sisällöntuotannon tehostamisen haasteet. Ei ole itsestään selvää, että korkeakoulukontekstissa web-pohjaisen sosiaalisen verkostopalvelun käyttöaktiivisuus saavuttaisi heti palvelun alusta lähtien sellaisen tason, että se säilyisi itseohjautuvasti yhtä korkeana koko lukuvuoden ajan. Itse asiassa, vapaaehtoiseen toimintaan perustuvan palvelun käyttömotivaatio vaatii palvelulta erityisen merkittävää lisäarvoa ((Silius et al., 2010a)).

Tutkimus-casessa keskityttiin tutkimaan onko TTY-Piirin käyttöaktiivisuus suoraan yhteydessä sosiaaliseen verkostoitumiseen, ts. verkostoituvatko palvelun käyt-

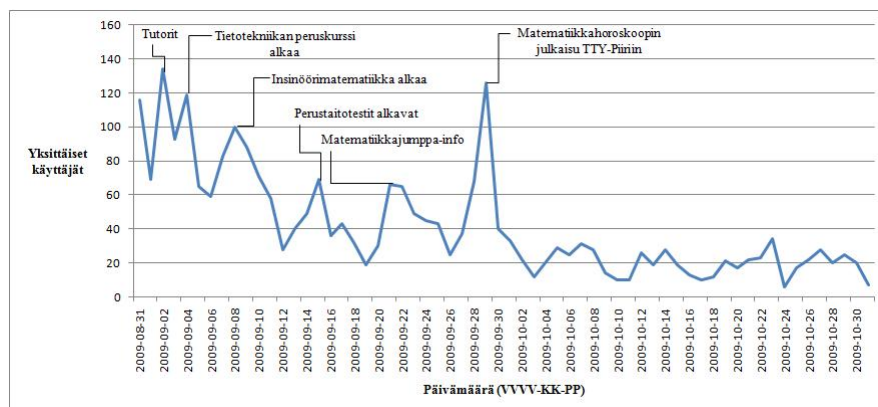
täjät toistensa kanssa enemmän, kun palvelun käyttöaste kasvaa. Kun palveluun tehtiin jokin interventio, pystyttiin käytönseurannalla havaitsemaan selvästi käyttöaktiivisuuden kasvu, mutta sosiaalisen verkostoitumisen aktiivisuuden tutkimiseen tarvittiin SNA-menetelmiä.

4.3 Interventiot ja käytönseuranta

Syksyllä 2009 TTY-Piiri julkaistiin opintonsa aloittaville uusille opiskelijoille. Jo ennen syksyä palvelu oli opiskelijoiden käytössä ja uusien opiskelijoiden lisäksi palvelua oli markkinoitu syksyn 2009 opiskelijatutoreille. Lukuvuoden alettua syyskuun alussa palvelun käyttöastetta seurattiin käyttäen *Google Analytics* -työkalua. Samalla palveluun tehtiin interventioita, joilla pyrittiin vaikuttamaan TTY-Piirin käyttöaktiivisuuteen. Nämä interventiot olivat lähinnä palvelun markkinointia, mutta myös uusien toiminnallisuuksien julkaisuja.

Käytönseurannan avulla pystyttiin selvästi havaitsemaan, että interventiot olivat suoraan yhteydessä palvelun käytön aktivoitumiseen. Selvästikin aina, kun TTY-Piiriä mainostettiin opetuksen yhteydessä, käyttäjämäärät lähtivät nousuun. Edelleen, esim. uuden toiminnallisuuden, nk. Matematiikkahoroskoopin, julkaisu, nosti käyttöaktiivisuutta merkittävästi, mutta vain hetkellisesti.


Kuvassa 4.2 on esitetty yksittäisten käyttäjien päivittäiset määrät TTY-Piirissä. Lisäksi kuvaan on merkitty tiettyjen interventioiden ajankohdat. Huomattavaa on, kuinka selvästi interventiot ovat lisänneet juuri sen hetkistä käyttöaktiivisuutta, mutta interventioiden pitkäjänteinen vaikutus käyttöön on ollut vähäistä. Kiinnostavaa olikin tietää, oliko interventioiden aiheuttama hetkellinen käytön lisäys vaikuttanut palvelun sisäiseen verkostoitumiseen. Tätä tutkittiin SNA-menetelmin.



Kuva 4.2: TTY-Piirin käyttöaktiivisuus syksyllä 2009.

4.4 Verkostodata

Luvussa 3.2 esiteltyä prosessia käytettiin TTY-Piirin SNA-analyysin toteutuksessa. TTY-Piiriin on toteutettu automaattinen lokidatatyökalu, joka tulostaa pyydetyn lokidatan XML-muodossa (ks. kuva 4.3). Edelleen tämä data voidaan syöttää Wille-ympäristöön, jolla siitä muokataan SNA-ohjelmistojen kanssa yhteensopivaa syötedataa.



Kuva 4.3: TTY-Piirin automaattinen lokidatatyökalu.

TTY-Piirin lokidatan tuontityökalulla tuotiin palvelun tietokannoista XML-formaatissa tieto palvelun sisällä tehdyistä kaveripyynnöistä. Datasta suljettiin pois tutkijoita ja palvelun kehittäjiä koskeneet tiedot ja tapahtumat. Tämä data muunnettiin edelleen Pajek-ohjelman syöteformaattiin sekä Matlab-ohjelman matriisi-syötteeksi. Data kerättiin ajalta 31.8.2009 - 31.10.2009. Edelleen data lohkottiin kolmeen osaan siten, että verkostolle voitiin määrittää alkutila (31.8.2009), välitila (30.9.2009) sekä lopputila (31.10.2009). Täten pyrittiin tutkimaan (ja visualisoimaan) verkoston kehitystä tutkittavalla aikavälillä.

4.4.1 XML-formaatti

TTY-Piirin lokidata XML-formaatissa pitää sisällään enemmän tietoa, kuin siitä datamuunnoksessa tuotetut formaatit. Käytännössä muunnoksessa tehdään tulkintaa sen mukaan, mitä halutaan analysoida ja miten. XML tarjoaa hyvät mahdollisuudet datan muuntamiselle useisiin eri formaatteihin, tosin tämä työ vaatii oikeanlaiset välineet. XML-muunnosten tekeminen käsin suurelle verkostodatalle vaatii yksinkertaisesti liikaa aikaa ja vaivaa.

Liitteessä B on esitetty osa TTY-Piirin kaveripyynnöistä kerätystä lokidatasta

ajalta 31.8.2009 - 31.10.2009 XML-formaatissa. Kokonaisuutena data koostui 8510 rivistä XML-dataa, eikä sen esittäminen kokonaisuutena ole järkevää. Alla on kuitenkin esitetty datasta suoritettun analyysin kannalta olennaisimmat kohdat.

Lokidatan alussa luetellaan verkoston toimijat (*actorlist*) ja osa heidän palveluun syöttämistään henkilö- ja kuvaustiedoista. Jokaisella toimijalla on oma tunnus (*actor id*), jonka avulla käyttäjät yksilöidään, mutta heidän yksityisyyttä ei loukata (esim. nimitietoja ei kerätä). Edelleen datamuunnoksessa kaikki muu yksilöä koskeva tieto jätetään datasta pois, sillä sen käytölle ei ole tarvetta kvantitatiivisissa analyysissä.

Toimijalistauksen jälkeen seuraa tapahtumalistaus (*eventlist*), jossa luetellaan edellä määritellyn toimijajoukon välillä tapahtuneet toiminnot. Edellä olevassa datasta on siis keskitytty kaveripyyntöihin. Tällä toiminnalla on tekijä, eli kaveripyyntöä esittäjä (*requestee-uid*), sekä kohde, eli toimija, jolle kaveripyyntö esitetään (*target_uid*). Kun tekijä ja kohde ovat eri toimijoita, on kyseessä kaveripyyntö käyttäjien välillä. Jos tekijä ja kohde ovat samat, tarkoittaa se tässä tapauksessa kaveripyyntöä hyväksymistä palvelun sisällä.

Tapahtumilla on aikaleima (*timestamp*), jonka avulla tapahtumat voidaan järjestää kronologiseen järjestykseen. Ensimmäisillä tapahtumilla aikaleima osoittaa niiden sijoittuvan aikaisemmalle ajanjaksolle, kuin tutkimuksessa keskitytään. Tämä johtuu siitä, että aikaisemmillä tapahtumilla luodaan verkoston alkutila, johon verkoston kehitystä verrataan.

4.4.2 Pajek-formaatti

Edelleen XML-datasta voidaan tuottaa SNA-dataa sopivilla datamuunnoksilla. Wille-ympäristön avulla voidaan tuottaa edellä esitetystä datasta Pajek-ohjelmiston kanssa yhteensopivaa syötedataa .paj-tiedostoformaatissa (formaatin määrittely ks. Batagelj ja Mrvar (2010)).

Pajek-aikasarjaformaatti

Kuitenkin tutkimuksessa haluttiin tuottaa kolmivaiheinen aikasarja verkoston evoluutiosta tarkastelemalla verkoston sisällä tapahtuneita kaveripyyntöjä ja niiden hyväksymisiä. Koska tällaista Pajek-aikasarjamuunnosta ei oltu toteutettu Willeen, se tehtiin räätälöidysti käsin editoimalla XML-dataa sopivilla välineillä, kuten Excelillä. Samalla tehtiin tulkintaa verkoston toimijoiden välisistä suhteista. Verkostossa toimijoiden välillä on suuntaamaton dikotominen yhteys, jos XML-datasta löytyy tapahtumapari "kaveripyyntö - pyynnön hyväksyminen".

Liitteessä C on esitetty osa Pajek-datasta, sillä kaiken kaikkiaan syötetiedosto sisälsi 2358 riviä tietoa, ja täten sen tulostaminen kokonaisuudessaan ei ole järkevää. Pajekin aikasarja-formaatti on esitelty mm. teoksessa Batagelj ja Mrvar (2010). Alla

olevaan esitykseen on otettu mukaan olennaisimmat osat datasta.

Pajek-aikasarjadataan alussa määritetään verkosto kokonaisuudessaan. Verkostolle on määritetty nimi, jonka jälkeen luetellaan verkoston toimijat. Luettelossa on juokseva järjestysluku, jota seuraa toimijan XML-datasta periytyvä tunnus. Edelleen toimijoille on annettu kolme reaalitylukuparametria, jotka määrittävät toimijan solmun sijainnin graafissa. Nämä on tässä datassa alunperin kaikille toimijoille merkitty nolliksi, mutta Pajekin avulla niille on saatu nolasta poikkeavia arvoja, kun graafi on ladottu uudelleen tiettyä graafinladonta-algoritmia (Kamada-Kawai) käyttäen. Rivin lopussa on hakasuluissa merkitty ajanhetket, joilla toimija on verkostossa aktiivinen, eli ns. aktiivisuusalue. Mikäli hakasulkumerkintää ei ole, on toimijan aktiivisuusalue sama kuin edellisen toimijan aktiivisuusalue.

Toimijalistauksen jälkeen luetellaan toimijoiden väliset yhteydet se., toimijalistauksessa toimijan järjestysluku vastaa nyt toimijan id-tunnusta. Yhteys merkitään yhdelle riville kolmilukuisena yhdistelmänä, jossa ensimmäiset kaksi lukua ilmoittavat toimijaparin id-tunnukset ja kolmas viimeinen luku yhteyden painon, joka tässä dikotomisessa verkostossa on kaikilla yksi. Lisäksi rivin lopussa on hakasuluissa merkitty yhteyden aktiivisuusalue. Mikäli hakasulkumerkintää ei ole, on yhteyden aktiivisuusalue sama kuin edellisen yhteyden aktiivisuusalue.

Koska verkostolle määritettiin kolme eri tilaa, seuraa verkoston kokonaismäärittelyn jälkeen datassa kolme vastaavanlaista määrittelyä kullekin ajanhetkelle. Näissä on siis lueteltu vain ne toimijat ja yhteydet, jotka ovat aktiivisia ko. hetkellä. Tälle kumuloituvalle datalle tämä tarkoittaa sitä, että ensimmäisessä ajanhetkessä aktiivisia ovat ne toimijat ja niiden väliset yhteydet, jotka XML-lokidatasta nähdään tapahtuneiksi ennen ko. ajankohtaa. Seuraavassa ajankohdassa edellisten toimijoiden ja yhteyksien lisäksi tulevat aktiivisiksi ne toimijat ja yhteydet, jotka nähdään XML-datasta tapahtuneiksi ensimmäisen ajanhetken ja tämän kyseisen ajanhetken välillä. Seuraava ajanhetki määritellään taas vastaavalla tavalla tarkastelemalla XML-dataa ja kumulatiivisesti lisäämällä toimijoita ja yhteyksiä edelliseen ajanhetkeen verrattuna.

4.4.3 Matlab-formaatti

Matlab-formaatissa on yksinkertaisesti annettu kullekin ajanhetkelle sosiomatriisi **X**. Tämä sosiomatriisi on symmetrinen ja dikotominen neliömatriisi. Laskennallinen analyysi on suoraan tehtävissä aiemmin esitetyillä määritelmillä näiden sosiomatriisien avulla. Seuraavassa on osa eräästä XML-datasta muunnetusta 225×225 -sosiomatriisista.

X =

Columns 1 through 10

0	0	0	0	0	0	0	0	0	0	...
0	0	0	0	0	0	0	0	0	0	...
0	0	1	1	0	0	0	0	1	0	...
0	0	1	1	0	1	1	1	1	1	...
0	0	0	0	0	0	0	0	0	0	...
0	0	0	1	0	1	1	1	1	1	...
0	0	0	1	0	1	1	1	1	1	...
0	0	0	1	0	1	1	1	1	1	...
0	0	1	1	0	1	1	1	1	1	...
0	0	0	0	0	0	0	1	1	1	...
.
.
.
0	0	0	0	0	0	0	0	0	0	...
0	0	0	0	0	0	0	0	0	0	...
0	0	0	0	0	0	0	0	0	0	...
0	0	0	0	0	0	0	0	0	0	...
0	0	0	0	0	0	0	0	0	0	...

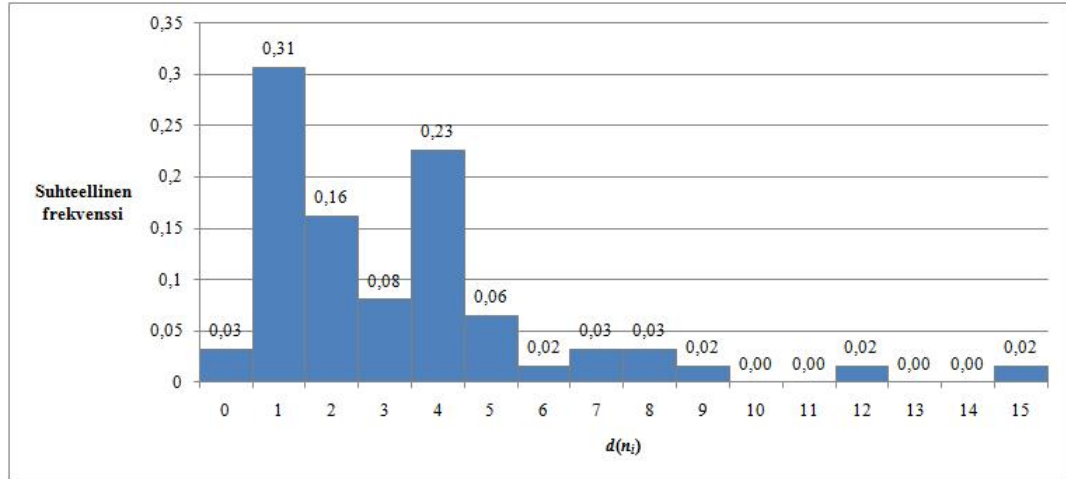
4.5 Analyysi

Verkoston analysoinnissa keskityttiin toimijoiden kohdalla tutkimaan näiden astelukuja, keskeisyysastetta, läheisyyttä ja välillisyyttä. Verkostoa kokonaisuutena tarkasteltiin sen tiheyden sekä toimijoiden keskittyneisyysasteen, läheisyyden keskittyneisyyden ja välillisyyden keskittyneisyyden kautta. Tällä tavoin pyrittiin vastaamaan edellä esitettyyn tutkimuskysymykseen se., jos verkostoa ja sen toimijoita kuvaavissa SNA-tunnusluvuissa on havaittavissa selvää kasvavaa muutosta interventioiden myötä (alkutila-välitila), voitaisiinko tällöin tulkita interventioiden vaikuttavan käyttöaktiivisuuden lisäksi myös sosiaaliseen verkostoitumiseen palvelun sisällä. Edelleen ajanjaksolla, jolla käyttöaktiivisuus on vähäisempää, eikä interventioita-kaan toteutettu (välitila-lopputila), ei tulisi olla suuria muutoksia verkoston tiheyteen tai toimijoiden keskeisyysasteisiin. Nämä analyysin tulokset on keskeisimmiltä osiltaan esitetty artikkelissa Silius et al. (2010a).

Alkutilan verkostolle, jossa oli toimijoita yhteensä 62 ja toimijoiden välisiä yhteyksiä yhteensä 101, saatiin määritettyä tiheydeksi $\Delta = 0.0457$. Tämän mukaan verkosto ei siis ole erityisen tiivis, mikä näkyy verkostossa mm. siinä, että se koostuu useista komponenteista.

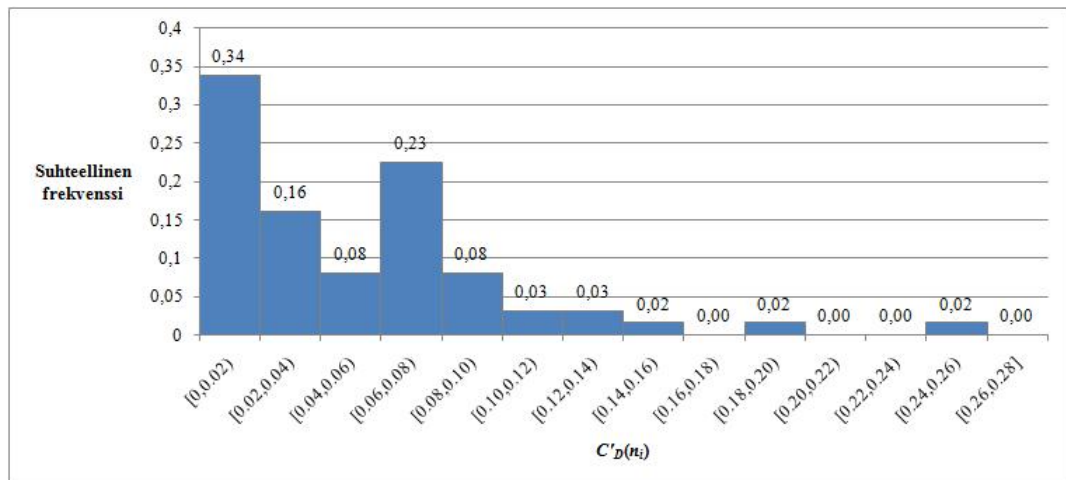
Toimijoiden astelukujen $d(n_i)$ jakauma on esitetty kuvassa 4.4. Toimijoiden keskiaste $\bar{d} = 3$ ja astelukujen varianssi $S_d^2 = 7.899524$. Jakaumasta nähdään, etteivät

asteluvut jakaudu erityisen tasaisesti verkoston toimijoiden kesken. Samaa asiaa indikoi myös varianssi. Kuitenkin pääpaino on pienissä asteluvuissa, mikä on ilmeistä verkon alkutilassa, jossa toiminta ja aktiivisuus on vähäistä.



Kuva 4.4: TTY-Piirin alkutilan toimijoiden astelukujen jakauma.

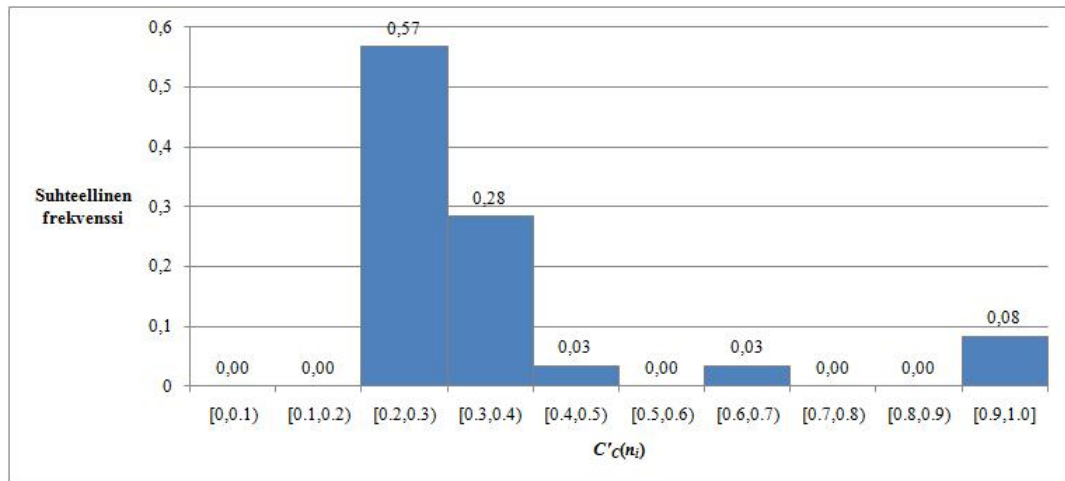
Alkutilan toimijoiden normeerattujen keskeisyysasteiden $C'_D(n_i)$ jakauma on esitetty kuvassa 4.5. Normeerattujen keskeisyysasteiden varianssiksi saatiin $S_D^2 = 0.002123$, mikä kertoo siitä, että verkostossa useimmat toimijat voidaan nähdä lähes yhtä keskeisinä. Verkoston keskittyneisyysaste $C_D = 0.002459$ kuitenkin indikoi sitä, että yleisesti verkoston toimijoiden keskeisyys ei ole merkittävällä tasolla. Verkostossa ei myöskään ole erityisen keskeisiä toimijoita, sillä toimijoiden joukosta korkeimmat keskeisyysasteet jäävät alle 0.26:n.



Kuva 4.5: TTY-Piirin alkutilan toimijoiden normeerattujen keskeisyysasteiden jakauma.

Verkoston alkutilassa oli yhteensä kaksi solmua, joiden asteluku oli nolla. Tarkasteltaessa läheisyyden ja välillisyyden tunnuslukuja, on nämä toimijat jätettävä pois tarkastelusta, sillä läheisyyden ja välillisyyden tarkastelussa tarvitaan solmujen välisien geodeesien määrittämistä. Näillä solmuilla geodeesit ovat määrittämättömiä, jolloin niillä ei ole vaikutusta muiden verkoston toimijoiden läheisyyksiin ja välillisyyksiin.

Kuvassa 4.6 on esitetty alkutilan verkoston toimijoiden normeerattujen läheisyyksien $C'_C(n_i)$ jakauma. Näiden varianssi $S_C^2 = 0.046058$ kertoo siitä, ettei läheisyyksissä ole koko verkoston tasolla suuria eroja toimijoiden välillä, vaikka kuvasta 4.6 voisi tulkita toisin. Huomattava on kuitenkin, että suurin osa toimijoiden normeeratuista läheisyyksistä asettuu välille $[0.2, 0.4]$. Läheisyyden keskittyneisyys $C_C = 0.024734$ puolestaan kertoo siitä, että verkostossa toimijat eivät ole erityisen läheisiä toistensa kanssa.

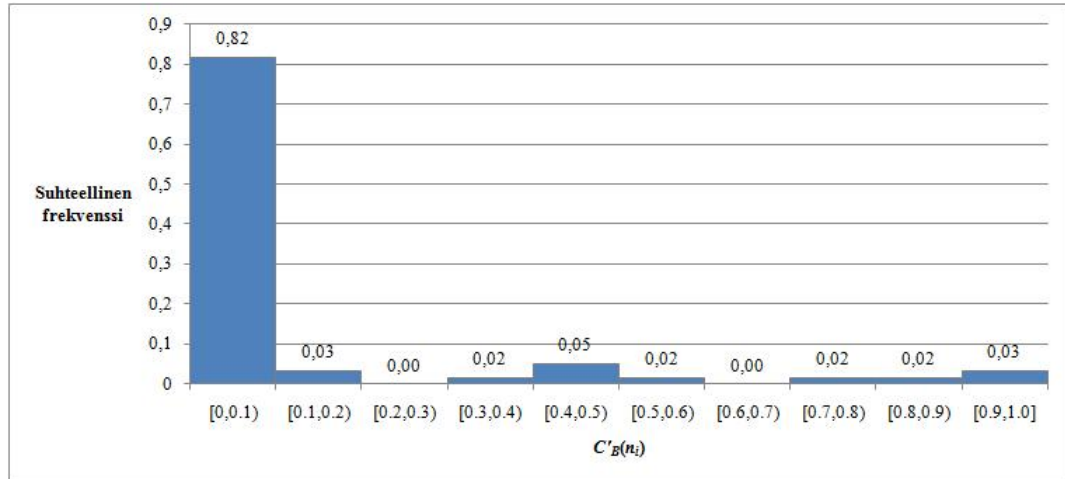


Kuva 4.6: TTY-Piirin alkutilan toimijoiden normeerattujen läheisyyksien jakauma.

Alkutilan toimijoiden normeerattuja välillisyyksiä $C'_B(n_i)$ tutkittaessa huomataan, että suuri osa toimijoiden välillisyyksistä asettuu alueelle $[0.00 - 0.01)$ (ks. kuva 4.7). Tämä kertoo yhdessä välillisyyden keskittyneisyyden $C_B = 0.010376$ kanssa siitä, etteivät verkoston toimijat keskimäärin ole suuresti osallisena monessa yhteydessä eri toimijaparien välillä. Normeeratun välillisyyden varianssi $S_B^2 = 0.059847$ kertoo, että normeeratuissa välillisyyksissä on pientä vaihtelua toimijoiden kesken.

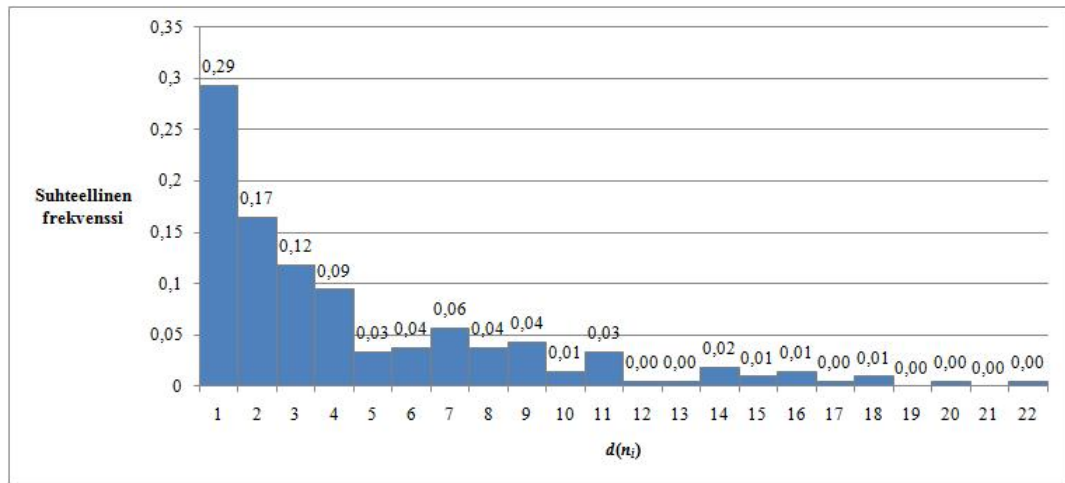
Välitilan verkoston, toimijoita yhteensä 212 ja näiden välisiä yhteyksiä yhteensä 480, tiheys $\Delta = 0.0213$ laski alkutilasta, mikä olennaisesti johtuu toimijamäärän merkittävästä kasvusta. Kuitenkin verkosto yhtenäistyi hieman, mikä näkyy pienempänä komponenttien määränä ja mm. siinä, että jokainen toimija on yhteydessä ainakin yhteen muuhun verkoston toimijaan.

Toimijoiden astelukujen $d(n_i)$ jakauma on esitetty kuvassa 4.8. Nyt keskiaste



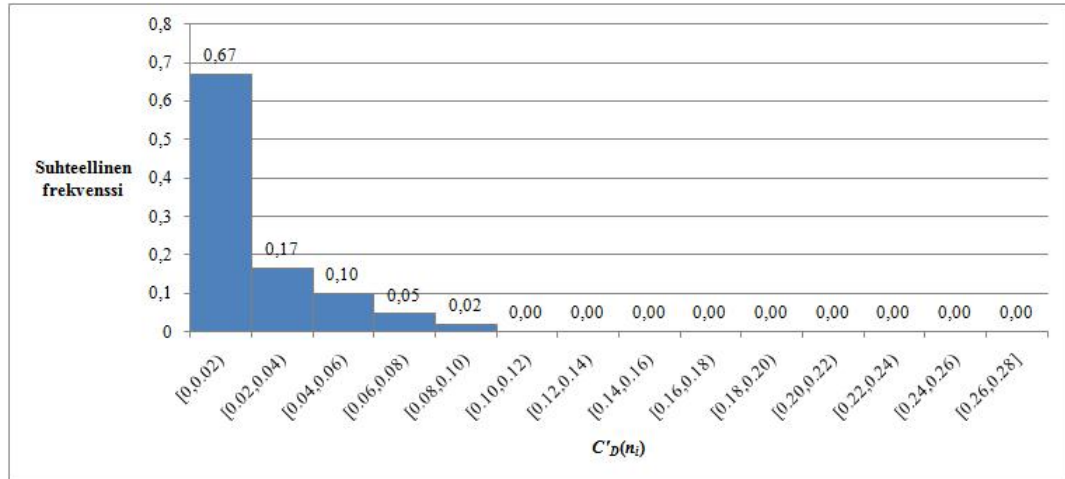
Kuva 4.7: TTY-Piirin alkutilan toimijoiden normeerattujen välillisyyksien jakauma.

$\bar{d} = 5$ ja astelukujen varianssi $S_d^2 = 18.83806$. Asteluvut asettuvat entistä enemmän alkupäähän. Verkostossa on kuitenkin toimijoita, joilla on suhteellisen suuri asteluku, mikä näkyy myös variassista.



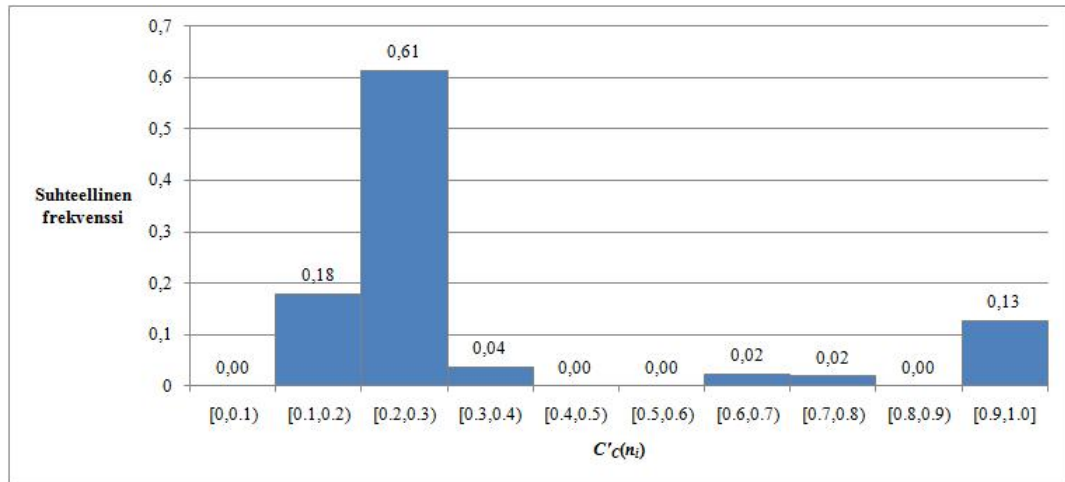
Kuva 4.8: TTY-Piirin välitilan toimijoiden astelukujen jakauma.

Välitilan toimijoiden normeerattujen keskeisyysasteiden $C'_D(n_i)$ jakauma on esitetty kuvassa 4.9. Normeeratun keskeisyysasteen varianssi $S_D^2 = 0.000375$ on alkutilaa huomattavasti pienempi, mikä kertoo siitä, että verkostossa yhä useammat toimijat ovat yhtä keskeisiä. Verkoston keskittyneisyysaste $C_D = 0.000316$ on myös alkutilan vastaavaa lukua pienempi, mikä kertoo siitä, että toimijoiden keskeisyys on yleisesti vain heikentynyt. Tähän syynä on merkittävä toimijamäärän kasvu, mikä ei ole ollut samanlaista verkoston toimijoiden välisten yhteyksien suhteen.



Kuva 4.9: TTY-Piirin välitilan toimijoiden normeerattujen keskeisyysasteiden jakauma.

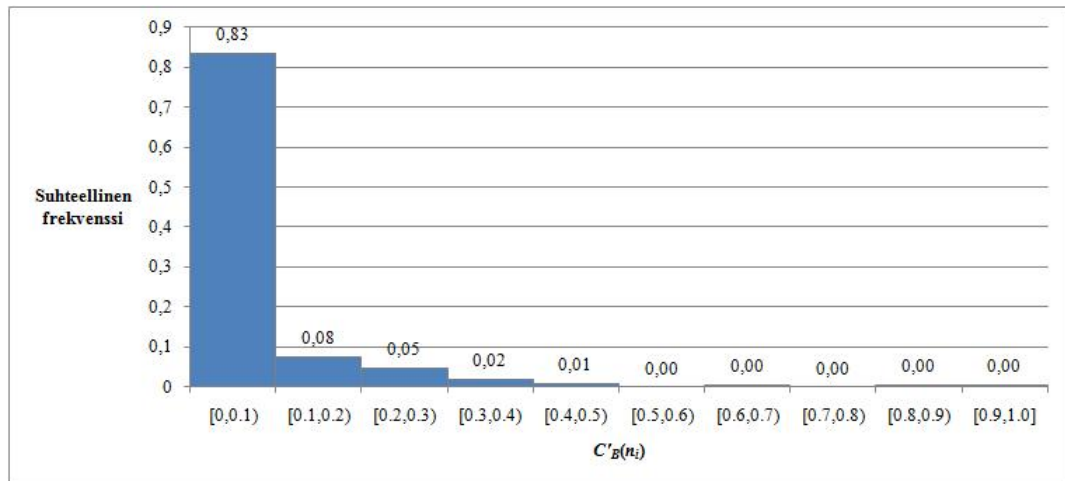
Välitilan verkoston toimijoiden normeerattujen läheisyyksien $C'_C(n_i)$ jakauma on esitetty kuvassa 4.10. Normeeratun läheisyyden varianssi $S_C^2 = 0.072363$ on kasvanut edellisestä, mikä näkyy myös hivenen leveämpänä jakaumana. Läheisyyden keskittyneisyys $C_C = 0.006669$ indikoi verkoston yleisen tason välillisyyden heikkenemistä. Ts. verkoston toimijat ovat verkostossa entistä etäämmällä toisistaan.



Kuva 4.10: TTY-Piirin välitilan toimijoiden normeerattujen läheisyyksien jakauma.

Välitilan toimijoiden normeeratut välillisyydet $C'_B(n_i)$ asettuvat edelleen pääosin alueelle $[0.00 - 0.01)$. Kuvan 4.11 jakaumasta nähdään edelleen, kuinka välillisyyden jakauma alkutilaan nähden hieman siirtyy vasemmalle painottuen enemmän pienille arvoille. Välillisyyden varianssi $S_B^2 = 0.016887$ ja välillisyyden keskittyneisyys $C_B = 0.003726$ indikoivat yhdessä sitä, että verkostossa toimijoiden osallisuus mui-

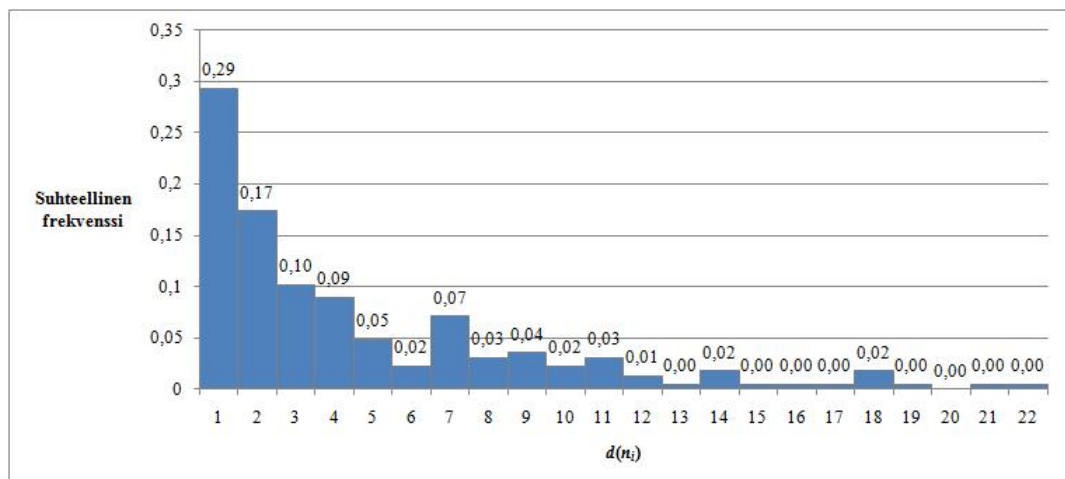
den toimijaparien välisiin yhteyksiin on heikentynyt. Myös tämä johtuu suurilta osin verkoston toimijoiden lukumäärän merkittävästä kasvusta.



Kuva 4.11: TTY-Piirin välitilan toimijoiden normeerattujen välillisyyksien jakauma.

Lopputilassa, toimijoita yhteensä 225 ja toimijoiden välisiä yhteyksiä yhteensä 516, ei ole tapahtunut merkittävää muutosta välitilaan. Verkoston tiheys ($\Delta = 0.0205$) on lähes sama kuin välitilalla. Myös muiden tunnuslukujen muutokset ovat vain nimellisiä.

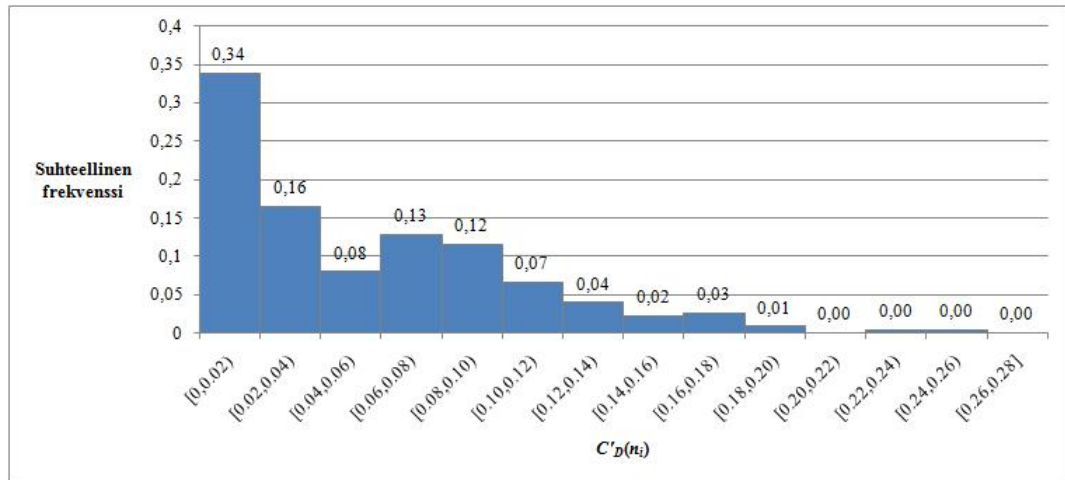
Kuvassa 4.12 on esitetty toimijoiden astelukujen $d(n_i)$ jakauma. Keskiaste on edelleen $\bar{d} = 5$. Astelukujen varianssi $S_d^2 = 19.79714$ on hivenen noussut. Jakaumasta nähdään, kuinka asteluvut asettuvat välitilan tapaan alkupäähän.



Kuva 4.12: TTY-Piirin lopputilan toimijoiden astelukujen jakauma.

Lopputilan toimijoiden normeerattujen keskeisyysasteiden $C'_D(n_i)$ jakauma on

esitetty kuvassa 4.13. Normeeratun keskeisyysasteen varianssi $S_C^2 = 0.002521$ on nyt puolestaan kasvanut välitilasta, mikä kertoo siitä, että jotkin välitilassa vähemmän verkostoituneet toimijat ovat nyt luoneet verkostossa uusia yhteyksiä hivenen enemmän. Verkoston keskittyneisyysaste $C_D = 0.00028$ on kuitenkin alkutilaa pienempi, mikä kertoo siitä, että toimijoiden keskeisyys verkostotasolla on edelleen hieman heikentynyt.

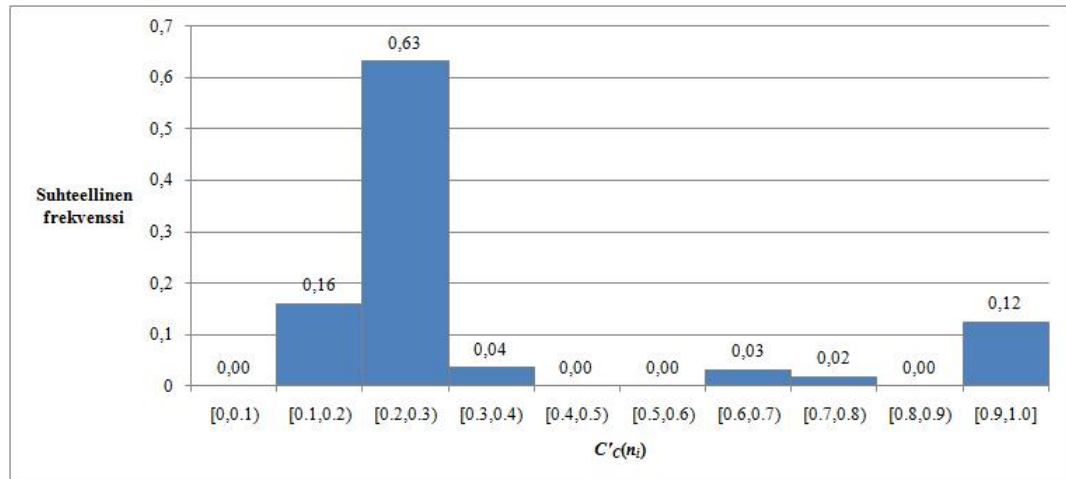


Kuva 4.13: TTY-Piirin lopputilan toimijoiden normeerattujen keskeisyysasteiden jakauma.

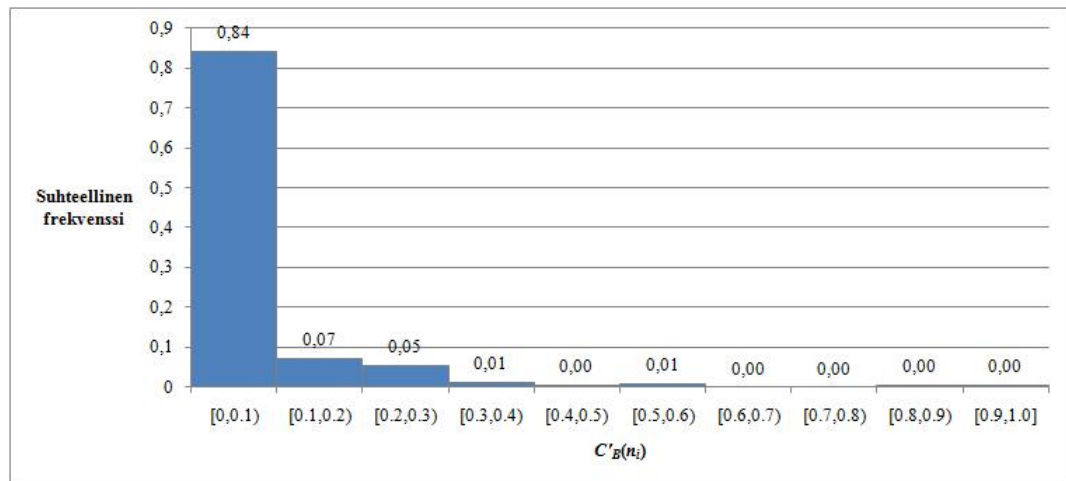
Verkoston toimijoiden normeerattujen läheisyyksien $C'_C(n_i)$ jakauma lopputilassa on esitetty kuvassa 4.14. Normeerattujen läheisyyksien varianssi $S_C^2 = 0.071279$ ei poikkea juuri lainkaan välitilasta. Sama pätee läheisyyden keskittyneisyydelle $C_C = 0,006317$, mikä kertoo yksinkertaisesti siitä, että verkoston toimijat eivät ole lähentyneet toisiaan merkittävästi välitilan ja lopputilan välillä.

Myös lopputilan toimijoiden normeeratut välillisyydet $C'_B(n_i)$ asettuvat suurilta osin alueelle $[0.00 - 0.01)$. Kuvan 4.15 jakaumasta nähdään kuitenkin, että verrattuna välitilaan, jakauman muilla alueilla on tapahtunut pientä kasvua. Normeeratun välillisyyden varianssi $S_B^2 = 0.016887$ on hivenen pienentynyt, mikä puoltaa osaltaan edellä mainittua jakauman tulkintaa. Välillisyyden keskittyneisyys $C_B = 0.003535$ on kuitenkin välitilaan nähden likimain yhtä suuri, mistä voidaan päätellä, etteivät verkoston muutokset ole vaikuttaneet toimijoiden osallisuuteen muiden toimijaparien välisissä yhteyksissä.

Liitteessä D on esitetty kootusti laskennallisen analyysin tulokset kullekin tilalle erikseen.



Kuva 4.14: TTY-Piirin lopputilan toimijoiden normeerattujen läheisyyksien jakauma.

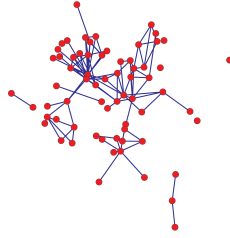


Kuva 4.15: TTY-Piirin lopputilan toimijoiden normeerattujen välillisyyksien jakauma.

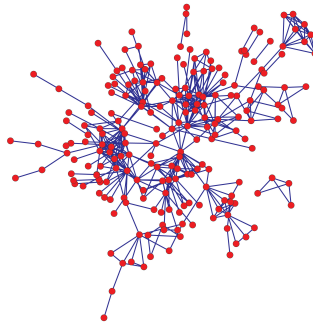
4.6 Visualisointi

Pajek-ohjelmiston avulla voidaan helposti visualisoida suuntaamattomina graafeina aikasarjan kolme eri tilaa. Nämä on esitetty kuvassa 4.16, jossa graafit on ladottu käyttäen Pajek-ohjelman Kamada-Kawai-graafinladontaalgoritmia.

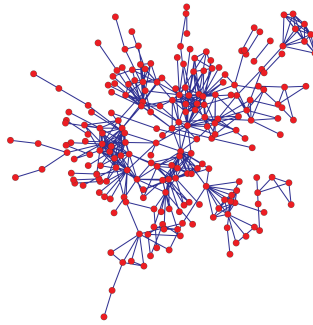
Aikasarjan visualisoinnista on helposti havaittavissa kuinka alkutilassa on useampia komponentteja kuin väli- ja lopputilassa. Verkoston evoluution kannalta olennainen huomio on se, että väli- ja lopputilalla ei ole visuaalisesti havaittavia merkityksellisiä eroja. Tämä tukee sitä selvää huomiota, jonka mukaan käyttöaktiivisuuden ollessa alhainen, ei sosiaalisessa verkostoitumisessakaan tapahdu merkittäviä muutoksia.



Alkutila (31.8.2009), $g=62$, $L=101$



Välitila (30.9.2009), $g=212$, $L=480$



Lopputila (31.10.2009), $g=225$, $L=516$

Kuva 4.16: TTY-Piirin aikasarjadatan visualisointi.

4.7 Tulokset

Laskennallisessa analyysissä saatujen tulosten mukaan interventioilla aikaansaatu käyttöaktiivisuuden lisäys tuotti verkostoon lisää toimijoita ja sitä kautta myös uusia

verkoston toimijoiden välisiä yhteyksiä. Kuitenkaan verkoston kasvaessa sen tiheys ei kasvanut, vaikka palvelun käytössä oli merkittäviä aktiivisuuspiikkejä.

Toimijoita kuvaavat tunnusluvut pysyivät suunnilleen samalla tasolla, mikä kertoo siitä, että verkostoon ajan kuluessa tulleet uudet toimijat verkostoituivat käytännössä reaalielämässäänkin tuttujen toimijoiden kanssa, mutta syvällisempi verkoston luominen jäi vähemmälle. Edelleen verkoston tunnusluvuissa ei voida sanoa tapahtuneen merkittäviä muutoksia. Tämä tarkoittaa sitä, että interventioilla saatiin siis aikaan lähinnä vain uusien toimijoiden liittäminen osaksi olemassa olevaa verkostoa, mutta aikaisempien toimijoiden aktiivisuutta verkostoitumisessa se ei suuremmin lisännyt.

Lisäksi voidaan sanoa, että koska verkoston käyttöaktiivisuus oli erityisesti väli- ja lopputilan välisenä ajanjaksona vähäistä, ei myöskään verkostossa tapahtunut merkittäviä muutoksia. Tämä on tietysti ilmeistä, sillä mikäli verkostopalvelussa ei ole käyttäjiä, ei myöskään verkostoitumisen välineitä ja toimintoja käytetä.

Vastauksena tutkimuskysymykseen voidaan sanoa, että tutkimus-casen tapauksessa käyttöaktiivisuuden kasvu on yhteydessä uusien toimijoiden verkostoitumiseen sekä verkostossa jo aiemmin olevien, että muiden uusien toimijoiden välillä. Käyttöaktiivisuuden kasvu ei kuitenkaan itsessään indikoi yleisen verkostoitumisen kasvua, mutta koska myös verkostoituminen palvelun sisällä on osa käyttöaktiivisuutta, on aktiivisuudella ilmeinen yhteys myös verkostoitumisasteeseen. Käyttöaktiivisuuden avulla ei kuitenkaan voida luotettavasti ennustaa verkostoitumisen tasoa.

4.8 Johtopäätökset ja pohdintaa

Sosiaalinen verkostoituminen ei ole tietenkään web-pohjaisessakaan verkostoitumispalvelussa pelkästään kaveriverkoston luomista, vaan analyysissä täytyisi myös huomioida toimijoiden väliset muut interaktiot, kuten keskustelut, yhteisesti tuotetut sisällöt jne. Tämä kuitenkin luo haasteen tulkinnalle ja vaatii kvantitatiivisten ja kvalitatiivisten tutkimusmenetelmien sitomista yhteen sopivaksi kokonaisuudeksi. Tässä tapauksessa ei kuitenkaan ensisijaisesti pyrittykään saamaan aukotonta vastausta tutkimuskysymykseen, vaan pikemminkin testaamaan SNA:n mahdollisuuksia tämän tyyppisessä tutkimuksessa. Voidaan todeta, että SNA:n määrälliset menetelmät tarjoavat mahdollisuuksia hyvinkin yksityiskohtaisen tarkastelun tekemiselle, kun käytettävissä olevaa dataa tulkitaan ja analysoidaan johdonmukaisesti oikein.

Merkittävä tulevaisuuden haaste on tutkia sosiaalisia verkostoja dynaamisina ajassa kehittyvinä kokonaisuuksina. Aikasarjat korostavat aikaa merkityksellisenä tekijänä SNA:ssa. Carley (2003) ottaa esiin uuden käsitteen DNA (*Dynamic Network Analysis*). Verkoston mallintaminen ajan mukaan on hyvin riippuvaista tulkinnasta. Aikasarjoissa joudutaan keräämään tieto aina diskreeteistä pisteistä ja jatkuvan mallin tuottaminen on hankalaa. Tarve olisikin analysoida verkostoja dynaamisi-

na järjestelminä, mutta haasteena on löytää tai määritellä niitä DNA:n metodeita, joilla kyetään suurten dynaamisten, monimoodisien sekä erityyppisiä yhteyksiä käsittävien verkostojen analysointiin.

5. YHTEENVETO

Tässä työssä on esitetty sosiaalisen verkostanalyysin matemaattisia menetelmiä web-pohjaisille sosiaalisille verkostoille. Menetelmät ovat monilta osin sovellettavissa hyvinkin erityyppisille verkostoille, joille lähtökohdittain voidaan tehdä samantyyppistä tulkintaa verkoston toimijoista ja niiden välisistä suhteista. SNA pyrkii kuitenkin ensisijaisesti keskittymään ihmisverkostoihin ja analysoimaan niitä näkökulmasta, joka ottaa huomioon myös ihmisten välisen toiminnan ja sen sisällön. Tämä korostaa tulkinnan merkitystä analyysin jokaisessa vaiheessa ja tuo samalla mukaan kvalitatiivisen verkostotutkimuksen. SNA-menetelmillä saatava määrällinen tutkimustieto ei itsessään aina kerro mitään, vaan SNA:n rinnalla tulee toteuttaa laadullista sisällön tutkimusta.

SNA:n matemaattiset menetelmät pohjautuvat pääosin graafiteoriaan ja laskennallinen analyysi hyödyntää matriisilaskennan menetelmiä. Sekä graafiteorian että matriisianalyysin mallit ovat sovellettavissa periaatteellisesti verkkoanalyysin jokaisella alueella, ei siis pelkästään sosiaalisille ihmisverkostoille. Tässä työssä esitetyt menetelmät on sovellettu vastaamaan SNA:n tarpeisiin, mutta monet ovat hyödyllisiä sellaisenaan myös monilla muilla verkkoteorian osa-alueilla. Edelleen, merkityksellistä on, kuinka malli määritellään ja mihin tutkimuskysymyksiin toteutettavalla analyysillä pyritään vastaamaan. Tulkinnan merkitystä ei voida tässäkään kohtaa aliarvioida.

Menetelmiä on case-tutkimuksena testattu analysoimalla Tampereen teknillisen yliopiston omaa sisäistä sosiaalista verkostopalvelua TTY-Piiriä. Analyysissä ei ole käytetty kaikkia esiteltyjä metodeita, mutta niiden hyödyntämiselle on sovelluskohteita ko. palvelun SNA-jatkotutkimuksessa. Osana case-tutkimusta on esitetty huomioita ja ratkaisuja yleisellä tasolla web-pohjaisten sosiaalisten verkostopalveluiden analyysissä, tärkeimpänä yleinen looginen prosessi, jolla verkkopalvelun lokidatasta voidaan kehittää SNA:lle validia verkostodataa ja edelleen kuinka analyysivaihe suhteutetaan osaksi tätä prosessia. Haasteena on luoda järjestelmä, joka tekee tämän prosessin automatisoiduksi.

LÄHTEET

- Batagelj, V., & Mrvar, A. (1998). Pajek: A Program for Large Network Analysis. *Connections* 21,2, 47-57.
- Batagelj, V., & Mrvar, A. (2010). *Pajek: Program for Analysis and Visualization of Large Networks*. Ljubljana.
- boyd d. m., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Meditated Communication*, 13(1).
- Carley, K. M. (2003). Dynamic Network Analysis. Teoksessa National Research Council: Breiger, R., Carley, K. M. & Pattison, P. (toim.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. Washington, DC.: The National Academies Press.
- Carrington, P. J., Scott, J., & Wasserman, S. (toim.). (2007). *Models and Methods in Social Network Analysis*. New York: Cambridge University Press.
- Denzin, N. K., & Lincoln, Y. S. (toim.). (2005). *The SAGE handbook of qualitative research*. Thousand Oaks, CA: Sage Publications, Inc.
- Eskola, J., & Suoranta, J. (2000). *Johdatus laadulliseen tutkimukseen*. Tampere: Osuuskunta Vastapaino.
- Freeman, L. C. (2009). Methods of Social Network Visualization. Teoksessa R. A. Meyers (toim.), *Encyclopedia of Complexity and Systems Science*. Berlin: Springer.
- Hanneman, R. A., & Riddle, M. (2005). *Introduction to Social Network Methods*. Riverside, CA: University of California. Saatavilla <http://faculty.ucr.edu/hanneman/nettext/>
- Huhtamäki, J., Nykänen, O., & Salonen, J. (2009a). Catalysing the Development of a Conference Workspace. *HCI International 2009 Conference Proceedings*, 5621/2009, 43-52.
- Huhtamäki, J., Nykänen, O., & Salonen, J. (toim.). (2009b). *Deliverable D10.12: Component-based visualisation system with collaborative OKS core scenarios*. OPAALS project deliverable, Phase 2.
- Johanson, J.-E., Mattila, M., & Uusikylä, P. (1998). *Johdatus verkostoanalyysiin*. Helsinki: Kuluttajatutkimuskeskus. Saatavilla <http://www.valt.helsinki.fi/vol/kirja/>
- Katz, L. (1953). A New Status Index Derived From Sociometric Analysis. *Psychometrika*, Vol. 18, No. 1, 39-43.
- Knoke, D., & Yang, S. (2008). *Social Network Analysis. Second Edition*. Los Angeles: Sage Publications.
- Metsämuuronen, J. (2003). *Tutkimuksen tekemisen perusteet ihmistieteissä. 2. uudistettu painos*. Helsinki: International Methelp Ky.

- Moreno, J. L. (1953). *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. Beacon, N.Y.: Beacon House Inc.
- Nooy, W. de, Mrvar, A., & Batagelj, V. (2005). *Exploratory Social Network Analysis with Pajek*. Cambridge, UK: Cambridge University Press.
- Nykänen, O., Salonen, J., Haapaniemi, M., & Huhtamäki, J. (2008). A Visualisation System for a Peer-to-Peer Information Space. *Proceedings of OPAALS 2008, 7-8 October 2008, Tampere, Finland*, 76-86.
- O'Reilly, T. (2005). *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. Saatavilla <http://oreilly.com/web2/archive/what-is-web-20.html>
- Preece, J., Maloney-Krichmar, D., & Abras, C. (2003). History and emergence of online communities. Teoksessa B. Wellman (toim.), *Encyclopedia of Community*.
- Rice, R. E., Borgman, C. L., Bednarski, D., & Hart, P. J. (1989). Journal-to-journal citation data: Issues of validity and reliability. *Scientometrics*, 15.3-4, 257-282.
- Ruohonen, K. (2006). *Graafiteoria*. Tampere: Tampereen teknillisen yliopiston opetusmoniste no. 5, uusi sarja.
- Scott, J. (2000). *Social Network Analysis. A Handbook. Second Edition*. London: Sage Publications.
- Silius, K., Miilumäki, T., Huhtamäki, J., Tebest, T., Meriläinen, J., & Pohjolainen, S. (2010a). Social Media Enhanced Studying and Learning in Higher Education. *IEEE EDUCON Education Engineering 2010 - The Future of Global Learning Engineering Education*.
- Silius, K., Miilumäki, T., Huhtamäki, J., Tebest, T., Meriläinen, J., & Pohjolainen, S. (2010b). Students' Motivations for Social Media Enhanced Studying and Learning. *Knowledge Management & E-Learning: An International Journal (KM&EL), the Special Issue on Technology Enhanced Learning, Vol. 2, No. 1*, 51-67.
- Tampubolon, G. (2004). *Introduction to Social Network Analysis*. Manchester: CCSR. Saatavilla <http://www.ccsr.ac.uk/methods/publications/snacourse/snacourseweb.html>
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

A. WARSHALLIN ALGORITMI: MATLAB-KOODI

```
% Name:      Reachability, Warshall Algorithm
% Info:      Evaluates the reachabilities of the nodes in
%            the valued directed graph
% Author:    Thumas Miilumäki
% Organisation: Department of Mathematics,
%            Tampere University of Technology, Finland

function [R] = warshall(X)
g=size(X);
g=g(1,1);
a=1;
E=zeros(g,g,g);
for i=1:g
    for j=1:g
        if (X(i,j)~=0)
            E(i,j,a)=1;
        else
            E(i,j,a)=0;
        end
    end
end
for b=1:g
    for i=1:g
        for j=1:g
            if (E(j,i,a)==1)
                for k=1:g
                    E(j,k,a+1)=max(E(j,k,a),E(i,k,a));
                end
            end
        end
    end
end
a=a+1;
end
R=E(:, :, a);
```

B. TTY-PIIRIN LOKIDATA XML-FORMAATISSA

```
<?xml version="1.0" encoding="utf-8"?>
<logdata service="www.tut.fi/piiri">
<actorlist>
<actor id="0">
  <uid>35</uid>
  <horoskooppi></horoskooppi>
  <perustaitotesti></perustaitotesti>
  <kotipaikka>Kajaani</kotipaikka>
  <ika>01-01-1970</ika>
  <sukupuoli></sukupuoli>
  <koulutusohjelma>Tietotekniikka</koulutusohjelma>
</actor>
<actor id="1">
  <uid>161</uid>
  <horoskooppi></horoskooppi>
  <perustaitotesti></perustaitotesti>
  <kotipaikka></kotipaikka>
  <ika>01-01-1970</ika>
  <sukupuoli>mies</sukupuoli>
  <koulutusohjelma>Tietotekniikka</koulutusohjelma>
</actor>
<actor id="2">
  <uid>153</uid>
  <horoskooppi></horoskooppi>
  <perustaitotesti></perustaitotesti>
  <kotipaikka></kotipaikka>
  <ika>01-01-1970</ika>
  <sukupuoli>nainen</sukupuoli>
  <koulutusohjelma>Tietotekniikka</koulutusohjelma>
</actor>
<actor id="3">
  <uid>160</uid>
  <horoskooppi></horoskooppi>
  <perustaitotesti></perustaitotesti>
  <kotipaikka></kotipaikka>
  <ika>01-01-1970</ika>
  <sukupuoli>nainen</sukupuoli>
```

```
<koulutusohjelma>Tietotekniikka</koulutusohjelma>
</actor>

:

</actorlist>

<eventlist>
<event id="0">
  <aid>696</aid>
  <uid>35</uid>
  <target_uid>35</target_uid>
  <module>user_relationshipsactivity</module>
  <type>kaveri</type>
  <operation>approved</operation>
  <data>
    <requestee-uid>103</requestee-uid>
    <relationship-type-id>2</relationship-type-id>
  </data>
  <timestamp>2009-07-10 15:38:53</timestamp>
  <unixtime>1247229533</unixtime>
</event>
<event id="1">
  <aid>801</aid>
  <uid>161</uid>
  <target_uid>160</target_uid>
  <module>user_relationshipsactivity</module>
  <type>kaveri</type>
  <operation>approved</operation>
  <data>
    <requestee-uid>160</requestee-uid>
    <relationship-type-id>2</relationship-type-id>
  </data>
  <timestamp>2009-08-26 12:32:17</timestamp>
  <unixtime>1251279137</unixtime>
</event>
<event id="2">
  <aid>802</aid>
  <uid>153</uid>
  <target_uid>4</target_uid>
  <module>user_relationshipsactivity</module>
  <type>kaveri</type>
  <operation>approved</operation>
  <data>
    <requestee-uid>4</requestee-uid>
    <relationship-type-id>2</relationship-type-id>
  </data>
  <timestamp>2009-08-26 12:32:27</timestamp>
```

```
<unixtime>1251279147</unixtime>
</event>
<event id="3">
  <aid>805</aid>
  <uid>153</uid>
  <target_uid>153</target_uid>
  <module>user_relationshipsactivity</module>
  <type>kaveri</type>
  <operation>approved</operation>
  <data>
    <requestee-uid>157</requestee-uid>
    <relationship-type-id>2</relationship-type-id>
  </data>
  <timestamp>2009-08-26 12:32:43</timestamp>
  <unixtime>1251279163</unixtime>
</event>

:

</eventlist>
</logdata>
```


C. TTY-PIIRIN SNA-DATA PAJEK-AIKASARJAFORMATISSA

```

*Network TTY-Piiri_-_Time_Serie
*Vertices 225
  1 "35"          0.4636    0.4886    0.5000 [1-*]
  2 "161"         0.3806    0.3286    0.5000
  3 "153"         0.3867    0.4220    0.5000
  4 "160"         0.4240    0.3729    0.5000
  5 "168"         0.1865    0.4683    0.5000
  6 "173"         0.3494    0.3780    0.5000
  7 "175"         0.3418    0.4052    0.5000
  8 "62"          0.3830    0.4325    0.5000
  9 "176"         0.3873    0.3722    0.5000
 10 "174"         0.3327    0.4881    0.5000

  :

223 "112"         0.1857    0.5306    0.5000 [2-*]
224 "229"         0.6891    0.6777    0.5000 [3-*]
225 "484"         0.1843    0.4751    0.5000

*Edges
  1 143 1 [1-*]
  2  4 1 [1-*]
  3 144 1 [1-*]
  3  45 1 [1-*]
  2 145 1 [1-*]
  2  3 1 [1-*]
  4 145 1 [1-*]
  3 146 1 [1-*]
  3  30 1 [1-*]
  5 147 1 [1-*]

  :

103 136 1 [3-*]
108 183 1 [3-*]
 60 225 1 [3-*]

*Network TTY-Piiri_-_Time_Serie in time 2009-08-31
*Vertices 67

```

1	"35"	0.4636	0.4886	0.5000 [1-*
2	"161"	0.3806	0.3286	0.5000
3	"153"	0.3867	0.4220	0.5000
4	"160"	0.4240	0.3729	0.5000
5	"168"	0.1865	0.4683	0.5000
6	"173"	0.3494	0.3780	0.5000
7	"175"	0.3418	0.4052	0.5000
8	"62"	0.3830	0.4325	0.5000
9	"176"	0.3873	0.3722	0.5000
10	"174"	0.3327	0.4881	0.5000

⋮

65	"154"	0.5865	0.3360	0.5000
66	"14"	0.6735	0.5370	0.5000
67	"28"	0.5769	0.4041	0.5000

*Edges

1	47	1	[1-*
2	4	1	[1-*
3	48	1	[1-*
3	41	1	[1-*
2	49	1	[1-*
2	3	1	[1-*
4	49	1	[1-*
3	50	1	[1-*
3	30	1	[1-*
5	51	1	[1-*

⋮

35	37	1	[1-*
10	35	1	[1-*
10	36	1	[1-*

*Network TTY-Piiri_-_Time_Serie in time 2009-09-30

*Vertices 213

1	"35"	0.4636	0.4886	0.5000 [1-*
2	"161"	0.3806	0.3286	0.5000
3	"153"	0.3867	0.4220	0.5000
4	"160"	0.4240	0.3729	0.5000
5	"168"	0.1865	0.4683	0.5000
6	"173"	0.3494	0.3780	0.5000
7	"175"	0.3418	0.4052	0.5000
8	"62"	0.3830	0.4325	0.5000
9	"176"	0.3873	0.3722	0.5000
10	"174"	0.3327	0.4881	0.5000

⋮

211	"386"	0.6532	0.2925	0.5000
212	"637"	0.7256	0.6108	0.5000
213	"112"	0.1857	0.5306	0.5000

*Edges

1	137	1	[1-*
2	4	1	[1-*
3	138	1	[1-*
3	45	1	[1-*
2	139	1	[1-*
2	3	1	[1-*
4	139	1	[1-*
3	140	1	[1-*
3	30	1	[1-*
5	141	1	[1-*
:			
40	53	1	[2-*
73	133	1	[2-*
51	67	1	[2-*

*Network TTY-Piiri_-_Time_Serie in time 2009-10-31

*Vertices 225

1	"35"	0.4636	0.4886	0.5000	[1-*
2	"161"	0.3806	0.3286	0.5000	
3	"153"	0.3867	0.4220	0.5000	
4	"160"	0.4240	0.3729	0.5000	
5	"168"	0.1865	0.4683	0.5000	
6	"173"	0.3494	0.3780	0.5000	
7	"175"	0.3418	0.4052	0.5000	
8	"62"	0.3830	0.4325	0.5000	
9	"176"	0.3873	0.3722	0.5000	
10	"174"	0.3327	0.4881	0.5000	
:					
223	"112"	0.1857	0.5306	0.5000	[2-*
224	"229"	0.6891	0.6777	0.5000	[3-*
225	"484"	0.1843	0.4751	0.5000	

*Edges

1	143	1	[1-*
2	4	1	[1-*
3	144	1	[1-*
3	45	1	[1-*
2	145	1	[1-*
2	3	1	[1-*
4	145	1	[1-*
3	146	1	[1-*

```
3   30 1 [1-*]
5   147 1 [1-*]

:

103 136 1 [3-*]
108 183 1 [3-*]
60  225 1 [3-*]
```

D. TTY-PIIRIN SNA:N TULOSTEN KOONTI

D.1 Alkutila: 31.8.2009

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
1	6	0.0984	2.6346	0.3878
2	4	0.0656	3.3846	0.0029
3	15	0.2459	2.4808	1.0000
4	4	0.0656	3.3846	0.0036
5	1	0.0164	1.0000	0.0000
6	4	0.0656	3.2692	0.0000
7	4	0.0656	3.2692	0.0000
8	12	0.1967	2.6731	0.4830
9	7	0.1148	3.2115	0.0268
10	5	0.0820	3.0769	0.5095
11	4	0.0656	3.0769	0.0326
12	4	0.0656	2.8077	0.0164
13	4	0.0656	3.0192	0.1046
14	8	0.1311	2.5769	0.9340
15	2	0.0328	3.3269	0.0000
16	1	0.0164	1.0000	0.0000
17	2	0.0328	4.1731	0.0000
18	2	0.0328	1.0000	0.0018
19	8	0.1311	3.9615	0.4424
20	5	0.0820	3.2115	0.7081
21	3	0.0492	4.0577	0.0000
22	2	0.0328	4.9231	0.0000
23	1	0.0164	4.3846	0.0000
24	4	0.0656	3.4038	0.0925
25	7	0.1148	2.8462	0.4092
26	3	0.0492	3.4231	0.0000
27	1	0.0164	4.9423	0.0000
28	2	0.0328	3.3269	0.0031
29	5	0.0820	2.9038	0.0566
30	4	0.0656	3.2500	0.0976

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
31	4	0.0656	3.7308	0.0888
32	3	0.0492	4.6731	0.0000
33	4	0.0656	3.7308	0.0888
34	3	0.0492	4.6731	0.0000
35	5	0.0820	3.9231	0.1804
36	4	0.0656	3.9423	0.0879
37	2	0.0328	4.8846	0.0000
38	2	0.0328	4.8846	0.0000
39	1	0.0164	4.9423	0.0000
40	9	0.1475	2.3846	0.8984
41	1	0.0164	3.4615	0.0000
42	0	0.0000		
43	1	0.0164	4.9038	0.0000
44	0	0.0000		
45	1	0.0164	4.9423	0.0000
46	1	0.0164	1.0000	0.0000
47	1	0.0164	3.6154	0.0000
48	1	0.0164	3.4615	0.0000
49	3	0.0492	3.4231	0.0000
50	1	0.0164	3.4615	0.0000
51	1	0.0164	1.0000	0.0000
52	1	0.0164	3.6538	0.0000
53	2	0.0328	3.4231	0.0000
54	2	0.0328	3.6346	0.0000
55	1	0.0164	4.0577	0.0000
56	1	0.0164	4.0000	0.0000
57	1	0.0164	1.5000	0.0000
58	2	0.0328	4.9231	0.0000
59	4	0.0656	4.0385	0.0063
60	4	0.0656	3.0769	0.0367
61	1	0.0164	1.5000	0.0000
62	1	0.0164	4.2308	0.0000
S^2	7.899524	0.002123	0.046058	0.059847
	$\bar{d} = 3$	$C_D = 0.002459$	$C_C = 0.024734$	$C_B = 0.010376$

D.2 Välitila: 30.9.2009

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
1	8	0.0379	3.2057	0.2138
2	4	0.0190	4.1371	0.0000
3	15	0.0711	3.3257	0.3242
4	10	0.0474	3.7600	0.0720
5	1	0.0047	1.0000	0.0000
6	4	0.0190	4.1486	0.0000
7	4	0.0190	4.1486	0.0000
8	16	0.0758	3.2743	0.3851
9	9	0.0427	4.0629	0.0056
10	14	0.0664	3.3543	0.4452
11	4	0.0190	3.8457	0.0086
12	4	0.0190	3.6171	0.0042
13	7	0.0332	3.5486	0.3992
14	20	0.0948	2.8686	1.0000
15	3	0.0142	3.5657	0.0404
16	2	0.0095	4.3371	0.0376
17	17	0.0806	2.9143	0.8383
18	2	0.0095	1.0000	0.0002
19	16	0.0758	3.5257	0.1104
20	11	0.0521	3.3029	0.1900
21	4	0.0190	3.8286	0.0000
22	7	0.0332	3.5714	0.0298
23	1	0.0047	4.4629	0.0000
24	11	0.0521	3.4686	0.1902
25	14	0.0664	3.5029	0.0568
26	10	0.0474	3.5429	0.0114
27	1	0.0047	4.5200	0.0000
28	2	0.0095	3.9943	0.0004
29	11	0.0521	3.3886	0.1456
30	5	0.0237	4.0800	0.0380
31	8	0.0379	4.2629	0.0124
32	4	0.0190	4.6457	0.0000
33	7	0.0332	3.7886	0.0542
34	5	0.0237	4.3714	0.0033
35	16	0.0758	3.6686	0.2531

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
36	13	0.0616	3.5200	0.1558
37	14	0.0664	3.5943	0.1804
38	18	0.0853	3.3600	0.3924
39	5	0.0237	4.4514	0.0011
40	18	0.0853	3.4743	0.2083
41	9	0.0427	3.8057	0.0286
42	10	0.0474	3.0914	0.2904
43	11	0.0521	3.4057	0.2630
44	1	0.0047	4.6057	0.0000
45	2	0.0095	4.3086	0.0004
46	1	0.0047	4.3543	0.0000
47	4	0.0190	4.1600	0.0028
48	11	0.0521	3.7371	0.1653
49	5	0.0237	4.0457	0.0472
50	9	0.0427	3.4114	0.0884
51	7	0.0332	4.6457	0.2438
52	9	0.0427	3.6514	0.0430
53	7	0.0332	3.8400	0.1151
54	11	0.0521	3.6629	0.0419
55	14	0.0664	3.7429	0.0727
56	1	0.0047	4.4800	0.0000
57	1	0.0047	1.0000	0.0000
58	2	0.0095	5.3600	0.0376
59	12	0.0569	3.4857	0.1940
60	9	0.0427	3.9657	0.0557
61	4	0.0190	4.5886	0.0020
62	3	0.0142	4.2229	0.0000
63	4	0.0190	5.6057	0.0749
64	1	0.0047	5.2514	0.0000
65	4	0.0190	1.0000	0.0003
66	2	0.0095	6.5886	0.0376
67	2	0.0095	5.6343	0.0000
68	1	0.0047	4.9600	0.0000
69	2	0.0095	4.3200	0.0055
70	3	0.0142	3.9657	0.1791
71	3	0.0142	1.0000	0.0001
72	3	0.0142	1.0000	0.0001
73	22	0.1043	3.4171	0.2917
74	2	0.0095	6.0057	0.0000
75	8	0.0379	3.6229	0.0377

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
76	2	0.0095	6.0057	0.0000
77	3	0.0142	4.4057	0.1195
78	1	0.0047	1.0000	0.0000
79	6	0.0284	3.0571	0.6095
80	1	0.0047	1.0000	0.0000
81	4	0.0190	4.2229	0.2308
82	6	0.0284	4.4057	0.0724
83	6	0.0284	4.6229	0.0125
84	1	0.0047	1.0000	0.0000
85	4	0.0190	4.6343	0.0120
86	3	0.0142	5.5886	0.0000
87	9	0.0427	4.6057	0.1244
88	6	0.0284	6.9771	0.0000
89	1	0.0047	1.0000	0.0000
90	7	0.0332	6.9714	0.0001
91	6	0.0284	6.9771	0.0000
92	7	0.0332	6.9714	0.0001
93	7	0.0332	6.9714	0.0001
94	8	0.0379	3.7429	0.0031
95	2	0.0095	4.8286	0.0000
96	5	0.0237	6.9829	0.0000
97	3	0.0142	5.0971	0.2891
98	8	0.0379	6.0114	0.2542
99	2	0.0095	5.6000	0.0376
100	2	0.0095	1.5000	0.0000
101	2	0.0095	1.5000	0.0000
102	3	0.0142	4.6171	0.1487
103	9	0.0427	3.6800	0.4812
104	8	0.0379	3.7257	0.0374
105	9	0.0427	3.6629	0.0465
106	5	0.0237	3.8114	0.0120
107	4	0.0190	4.6229	0.0000
108	15	0.0711	3.4114	0.1370
109	8	0.0379	3.5657	0.0015
110	11	0.0521	3.5943	0.0220
111	1	0.0047	1.0000	0.0000
112	2	0.0095	4.4057	0.0000
113	3	0.0142	3.7829	0.0000
114	5	0.0237	4.3886	0.1480
115	2	0.0095	1.5000	0.0000

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
116	4	0.0190	5.2629	0.0038
117	1	0.0047	4.6171	0.0000
118	3	0.0142	4.0571	0.0023
119	2	0.0095	4.6686	0.0000
120	6	0.0284	3.7600	0.0026
121	1	0.0047	1.0000	0.0000
122	1	0.0047	1.0000	0.0000
123	2	0.0095	5.6000	0.0376
124	3	0.0142	4.6400	0.0000
125	1	0.0047	1.0000	0.0000
126	2	0.0095	5.5943	0.0000
127	3	0.0142	4.3429	0.0611
128	6	0.0284	4.0457	0.1187
129	3	0.0142	5.1029	0.0104
130	3	0.0142	1.0000	0.0001
131	2	0.0095	1.3333	0.0000
132	2	0.0095	1.3333	0.0000
133	1	0.0047	4.4114	0.0000
134	1	0.0047	5.3314	0.0000
135	1	0.0047	4.7371	0.0000
136	1	0.0047	4.4114	0.0000
137	1	0.0047	4.2000	0.0000
138	1	0.0047	4.3200	0.0000
139	4	0.0190	4.1371	0.0000
140	1	0.0047	4.3200	0.0000
141	1	0.0047	1.0000	0.0000
142	1	0.0047	4.2686	0.0000
143	3	0.0142	3.8571	0.0111
144	3	0.0142	4.1771	0.0002
145	4	0.0190	3.9829	0.0002
146	1	0.0047	4.5429	0.0000
147	1	0.0047	1.5000	0.0000
148	6	0.0284	3.6514	0.0018
149	7	0.0332	3.6400	0.0013
150	8	0.0379	3.6114	0.0461
151	1	0.0047	1.5000	0.0000
152	1	0.0047	5.0743	0.0000
153	7	0.0332	3.7771	0.0072
154	9	0.0427	3.7314	0.0144
155	1	0.0047	4.3543	0.0000

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
156	3	0.0142	4.3543	0.0000
157	3	0.0142	3.9600	0.0304
158	2	0.0095	4.4343	0.0000
159	1	0.0047	4.2686	0.0000
160	1	0.0047	4.3543	0.0000
161	2	0.0095	4.8286	0.0000
162	1	0.0047	4.7314	0.0000
163	1	0.0047	1.0000	0.0000
164	2	0.0095	4.2571	0.0376
165	1	0.0047	4.8343	0.0000
166	3	0.0142	5.6171	0.0003
167	2	0.0095	5.1200	0.0000
168	1	0.0047	6.3543	0.0000
169	4	0.0190	1.0000	0.0003
170	1	0.0047	4.4800	0.0000
171	2	0.0095	4.2800	0.0098
172	2	0.0095	5.6229	0.0000
173	1	0.0047	7.5829	0.0000
174	2	0.0095	1.3333	0.0000
175	2	0.0095	1.3333	0.0000
176	1	0.0047	4.4114	0.0000
177	3	0.0142	5.0171	0.0748
178	2	0.0095	4.3771	0.0748
179	1	0.0047	5.0400	0.0000
180	1	0.0047	1.0000	0.0000
181	1	0.0047	4.0514	0.0000
182	1	0.0047	1.0000	0.0000
183	2	0.0095	4.1429	0.0000
184	1	0.0047	4.4686	0.0000
185	1	0.0047	4.5886	0.0000
186	2	0.0095	4.4571	0.0000
187	1	0.0047	1.0000	0.0000
188	1	0.0047	1.0000	0.0000
189	7	0.0332	6.9714	0.0001
190	1	0.0047	4.4000	0.0000
191	2	0.0095	5.5943	0.0000
192	3	0.0142	4.1257	0.0000
193	1	0.0047	6.5943	0.0000
194	4	0.0190	4.6343	0.0006
195	2	0.0095	4.4571	0.0000

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
196	7	0.0332	4.2514	0.0082
197	1	0.0047	4.4057	0.0000
198	3	0.0142	4.5600	0.0000
199	1	0.0047	1.0000	0.0000
200	2	0.0095	5.3943	0.0000
201	1	0.0047	4.6629	0.0000
202	1	0.0047	5.3829	0.0000
203	3	0.0142	3.7886	0.0022
204	4	0.0190	3.8514	0.0000
205	1	0.0047	4.4686	0.0000
206	1	0.0047	1.0000	0.0000
207	1	0.0047	1.0000	0.0000
208	1	0.0047	6.5943	0.0000
209	1	0.0047	1.0000	0.0000
210	1	0.0047	5.6000	0.0000
211	1	0.0047	5.3371	0.0000
212	3	0.0142	1.0000	0.0001
S^2	18.83806	0.000375	0.072363	0.016887
	$\bar{d} = 5$	$C_D = 0.000316$	$C_C = 0.006669$	$C_B = 0.003726$

D.3 Lopputila: 31.10.2009

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
1	8	0.0357	3.1838	0.2082
2	4	0.0179	4.1351	0.0000
3	15	0.0670	3.3459	0.2963
4	10	0.0446	3.7297	0.0776
5	1	0.0045	1.0000	0.0000
6	4	0.0179	4.1676	0.0000
7	4	0.0179	4.1676	0.0000
8	16	0.0714	3.2919	0.3528
9	9	0.0402	4.0649	0.0061
10	14	0.0625	3.4000	0.3661
11	4	0.0179	3.8595	0.0086
12	4	0.0179	3.5946	0.0041
13	8	0.0357	3.4919	0.4099
14	21	0.0938	2.8270	1.0000
15	4	0.0179	3.5189	0.0746
16	2	0.0089	4.1622	0.0129
17	18	0.0804	2.8811	0.8745
18	2	0.0089	1.0000	0.0002
19	17	0.0759	3.5189	0.1342
20	12	0.0536	3.2649	0.1819
21	4	0.0179	3.7892	0.0000
22	7	0.0313	3.5784	0.0249
23	1	0.0045	4.4108	0.0000
24	11	0.0491	3.4162	0.1923
25	14	0.0625	3.4378	0.0560
26	10	0.0446	3.4919	0.0105
27	1	0.0045	4.5135	0.0000
28	2	0.0089	3.9189	0.0008
29	11	0.0491	3.3405	0.1398
30	5	0.0223	4.0811	0.0369
31	8	0.0357	4.1027	0.0159
32	4	0.0179	4.5784	0.0000
33	7	0.0313	3.7514	0.0485
34	5	0.0223	4.3135	0.0031
35	18	0.0804	3.6541	0.2752

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
36	13	0.0580	3.5189	0.1259
37	14	0.0625	3.6378	0.1691
38	18	0.0804	3.3892	0.3792
39	5	0.0223	4.3892	0.0011
40	19	0.0848	3.4541	0.2090
41	10	0.0446	3.7838	0.0358
42	10	0.0446	3.0649	0.2835
43	11	0.0491	3.4378	0.2120
44	1	0.0045	4.5243	0.0000
45	2	0.0089	4.3297	0.0004
46	1	0.0045	4.3838	0.0000
47	4	0.0179	4.1622	0.0029
48	11	0.0491	3.7405	0.1687
49	5	0.0223	4.0000	0.0451
50	9	0.0402	3.4108	0.0759
51	7	0.0313	4.6703	0.2363
52	9	0.0402	3.6919	0.0403
53	8	0.0357	3.8216	0.1470
54	12	0.0536	3.6865	0.0534
55	14	0.0625	3.7838	0.0439
56	1	0.0045	4.4973	0.0000
57	1	0.0045	1.0000	0.0000
58	2	0.0089	5.3946	0.0366
59	12	0.0536	3.5027	0.1601
60	11	0.0491	3.8703	0.0819
61	4	0.0179	4.3730	0.0005
62	3	0.0134	4.2108	0.0000
63	4	0.0179	5.6324	0.0728
64	1	0.0045	5.2703	0.0000
65	4	0.0179	1.0000	0.0003
66	2	0.0089	6.6162	0.0366
67	2	0.0089	5.6595	0.0000
68	1	0.0045	4.8649	0.0000
69	3	0.0134	4.3081	0.0421
70	3	0.0134	4.0000	0.1712
71	3	0.0134	1.0000	0.0001
72	3	0.0134	1.0000	0.0001
73	22	0.0982	3.3568	0.2763
74	3	0.0134	4.7189	0.0366
75	8	0.0357	3.5622	0.0113

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
76	2	0.0089	4.7243	0.0000
77	3	0.0134	4.3892	0.1195
78	1	0.0045	1.0000	0.0000
79	6	0.0268	3.0703	0.5256
80	1	0.0045	1.0000	0.0000
81	5	0.0223	4.1459	0.2322
82	7	0.0313	4.3568	0.0629
83	7	0.0313	4.5784	0.0004
84	1	0.0045	1.0000	0.0000
85	4	0.0179	4.5946	0.0000
86	7	0.0313	4.5784	0.0366
87	10	0.0446	4.5622	0.1093
88	6	0.0268	6.9189	0.0000
89	1	0.0045	1.0000	0.0000
90	7	0.0313	6.9135	0.0001
91	6	0.0268	6.9189	0.0000
92	7	0.0313	6.9135	0.0001
93	7	0.0313	6.9135	0.0001
94	9	0.0402	3.7027	0.0028
95	3	0.0134	4.8054	0.0001
96	5	0.0223	6.9243	0.0000
97	3	0.0134	5.0324	0.2815
98	8	0.0357	5.9514	0.2476
99	2	0.0089	5.5892	0.0366
100	2	0.0089	1.5000	0.0000
101	2	0.0089	1.5000	0.0000
102	3	0.0134	4.6054	0.1446
103	11	0.0491	3.6378	0.5355
104	8	0.0357	3.6919	0.0344
105	9	0.0402	3.5946	0.0488
106	5	0.0223	3.7730	0.0107
107	4	0.0179	4.5568	0.0000
108	18	0.0804	3.2486	0.2859
109	9	0.0402	3.4973	0.0011
110	11	0.0491	3.5189	0.0221
111	2	0.0089	4.5135	0.0366
112	2	0.0089	4.3459	0.0000
113	3	0.0134	3.7459	0.0000
114	5	0.0223	4.3351	0.1255
115	2	0.0089	1.5000	0.0000

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
116	7	0.0313	4.3027	0.0411
117	2	0.0089	4.2378	0.0000
118	3	0.0134	4.0757	0.0023
119	2	0.0089	4.6270	0.0000
120	6	0.0268	3.7243	0.0026
121	1	0.0045	1.0000	0.0000
122	1	0.0045	1.0000	0.0000
123	2	0.0089	5.5892	0.0366
124	5	0.0223	4.5892	0.0000
125	1	0.0045	1.0000	0.0000
126	2	0.0089	5.5514	0.0000
127	5	0.0223	4.1838	0.0644
128	7	0.0313	3.7405	0.0985
129	3	0.0134	5.0270	0.0012
130	3	0.0134	1.0000	0.0001
131	2	0.0089	1.3333	0.0000
132	2	0.0089	1.3333	0.0000
133	1	0.0045	4.3514	0.0000
134	2	0.0089	4.1784	0.0110
135	2	0.0089	4.6378	0.0000
136	5	0.0223	4.5892	0.0000
137	2	0.0089	4.3027	0.0000
138	1	0.0045	5.3027	0.0000
139	4	0.0179	4.0162	0.0001
140	1	0.0045	1.0000	0.0000
141	1	0.0045	1.5000	0.0000
142	1	0.0045	1.5000	0.0000
143	3	0.0134	4.2919	0.0000
144	3	0.0134	3.9676	0.0285
145	4	0.0179	4.1351	0.0000
146	1	0.0045	4.3405	0.0000
147	1	0.0045	1.0000	0.0000
148	6	0.0268	3.6432	0.0016
149	7	0.0313	3.8162	0.0063
150	9	0.0402	3.7730	0.0132
151	1	0.0045	5.3514	0.0000
152	1	0.0045	4.4865	0.0000
153	7	0.0313	3.6324	0.0010
154	9	0.0402	3.5297	0.1177
155	1	0.0045	1.5000	0.0000

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
156	3	0.0134	4.1784	0.0003
157	3	0.0134	3.8162	0.0120
158	2	0.0089	4.1784	0.0000
159	1	0.0045	4.2865	0.0000
160	1	0.0045	1.5000	0.0000
161	2	0.0089	4.4108	0.0727
162	1	0.0045	4.1784	0.0000
163	1	0.0045	4.3405	0.0000
164	2	0.0089	4.4108	0.0000
165	1	0.0045	4.2865	0.0000
166	4	0.0179	3.7405	0.1621
167	2	0.0089	4.8108	0.0000
168	1	0.0045	4.7351	0.0000
169	4	0.0179	1.0000	0.0003
170	1	0.0045	4.8162	0.0000
171	2	0.0089	4.2757	0.0366
172	2	0.0089	1.3333	0.0000
173	1	0.0045	6.3892	0.0000
174	2	0.0089	5.1351	0.0000
175	2	0.0089	1.3333	0.0000
176	1	0.0045	4.4973	0.0000
177	3	0.0134	5.6432	0.0003
178	2	0.0089	5.6486	0.0000
179	1	0.0045	7.6108	0.0000
180	1	0.0045	1.0000	0.0000
181	1	0.0045	1.0000	0.0000
182	1	0.0045	4.3514	0.0000
183	2	0.0089	4.4378	0.0000
184	1	0.0045	4.3838	0.0000
185	1	0.0045	4.9946	0.0000
186	2	0.0089	4.2973	0.0183
187	1	0.0045	4.0649	0.0000
188	1	0.0045	1.0000	0.0000
189	7	0.0313	4.0973	0.0095
190	1	0.0045	4.4486	0.0000
191	2	0.0089	4.4378	0.0000
192	3	0.0134	3.7514	0.0022
193	1	0.0045	1.0000	0.0000
194	4	0.0179	4.6000	0.0005
195	3	0.0134	4.1459	0.0000

n_i	$d(n_i)$	$C'_D(n_i)$	$C'_C(n_i)$	$C'_B(n_i)$
196	7	0.0313	6.9135	0.0001
197	1	0.0045	1.0000	0.0000
198	7	0.0313	3.3946	0.0557
199	1	0.0045	6.5838	0.0000
200	2	0.0089	5.5514	0.0000
201	1	0.0045	4.6324	0.0000
202	1	0.0045	5.0757	0.0000
203	3	0.0134	4.5027	0.0000
204	4	0.0179	3.8162	0.0000
205	1	0.0045	5.5081	0.0000
206	1	0.0045	4.2432	0.0000
207	1	0.0045	4.6486	0.0000
208	1	0.0045	5.3297	0.0000
209	1	0.0045	4.3838	0.0000
210	2	0.0089	5.2865	0.0000
211	1	0.0045	4.4486	0.0000
212	3	0.0134	1.0000	0.0001
213	1	0.0045	1.0000	0.0000
214	1	0.0045	6.5838	0.0000
215	1	0.0045	1.0000	0.0000
216	1	0.0045	5.5568	0.0000
217	1	0.0045	5.1784	0.0000
218	1	0.0045	1.0000	0.0000
219	1	0.0045	4.4324	0.0000
220	2	0.0089	4.8108	0.0000
221	1	0.0045	5.7135	0.0000
222	1	0.0045	5.5730	0.0000
223	1	0.0045	1.0000	0.0000
224	2	0.0089	1.0000	0.0002
225	1	0.0045	4.8649	0.0000
S^2	19.79714	0.002521	0.071279	0.015764
	$\bar{d} = 5$	$C_D = 0.00028$	$C_C = 0.006317$	$C_B = 0.003535$